

Db 209 AVHTA-----TENIPOTPOKIMMMLPGVGVDEMTGVDGNTPLHADENYV---RYTPLE 261
197 --GGEGGSGPFTIDWT-----DNDFPDSRNG-RGDWTFDG 230
Db 262 ALDEESGDNEEPVEVEEPADEEVSVPISAIYEFENFENIWSIAHGWT-NG 320
QY 231 NRYDLT--DKNYSRDMILLATLRKQSESFNGVPRDDEPA 270
Db 321 QMFNATWYDSQITFSNIGMFAIDKE-----DDEEA 351

RESULT 2

Q45648 PRELIMINARY; PRT; 276 AA.
AC 045648;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE LICHENASE PRECURSOR.
GN BGAL.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1409;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=137;
RA MEDLINE=94288605; PubMed=7517127;
RA Taberner C., Coll P.M., Fernandez-Abalos J.M., Perez P.,
RA Santamaria R.I.;
RT "Cloning and DNA sequencing of bgaA, a gene encoding an endo-beta-1,3-
RT RT 1,4-glucanase, from an alkalophilic Bacillus strain (N137).";
RL Appl. Environ. Microbiol. 60:1213-1220(1994).
EMBL: Z12151; CA878135.1; -
DR HSSP: P23904; IAJK.
DR InterPro: IPR000757; -
DR pfam: PF00722; Glyco_hydro_16; 1.
DR PRINTS: PR00737; GLHYDRLASE16.
DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
KM Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 276 LICHENASE.
SQ SEQUENCE 276 AA; 31770 MW; D047F8A34CA9EBE2 CRC64;

Query Match 13.6%; Score 247.5; DB 2; Length 276;
Best Local Similarity 36.1%; Pred. No. 1.9e-11;
Matches 61; Conservative 20; Mismatches 75; Indels 13; Gaps 6;

QY 30 FSGAELTYLLEVOYQKFEARKMAAGTVSMFLYONGSEINDGRPWVEVDIEVGKNP 89
Db 96 YKAGELRTNDFYHGLFEVSMKPAKSTGVSSFTY-TGFWMDENDWDEIDIEFGKDT 154
QY 90 GSFQSNITTCGAKAQKTSKHNHVAVSPAADQAFHTYGLFEMTPNVRWTVDSQEVKTEGGQ 149
Db 155 TKQGFNYFNGVGG--NHHYHGLGDADDFVTYAFEMRPESIRNVRVNGELVHTA---- 207
QY 150 VSNULTGT-OGLRFNLMSSESA-AMVGOFDESKLPLOFIMWVYKYP 196
Db 208 TENIPOTPOKIMMMLPGVGVDEMTGVRNEDPVTQYDMV---KYTP 253

RESULT 3
Q45691 PRELIMINARY; PRT; 801 AA.
AC 045691;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMblrel. 16, Last annotation update)
DE FAMILY 11 XYLANASE /FAMILY 16 BETA (1,3-1,4) GLUCANASE.
GN XYND.
OS Ruminococcus flavefaciens.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Ruminococcus.
OX NCBI_TaxID=1265;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=17;
RA Aurilia V., Martin J.C., Munro C.A., Mercer D.K., Flint H.J.;
RT "Organisation and strain distribution of genes responsible for the
RT utilization of xylans by the rumen cellulolytic bacterium Ruminococcus
RT flavefaciens 17";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ132472; CAB51934.1; -
DR HSSP: P23904; IAJK.
DR InterPro: IPR000757; -
DR InterPro: IPR001137; -
DR InterPro: IPR003305; -
DR pfam: PF00457; Glyco_hydro_11; 1.
DR pfam: PF00722; Glyco_hydro_16; 1.
DR pfam: PF02018; CBD_6; 1.
DR PRINTS: PR00911; GLHYDRLASE11.
DR PRINTS: PR00737; GLHYDRLASE16.
DR PROSITE: PS00776; GLYCOSYL_HYDROL_F11_1; 1.
DR PROSITE: PS00777; GLYCOSYL_HYDROL_F11_2; 1.
DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
FT INIT_MET 0
FT SEQUENCE 801 AA; 88887 MW; 9F7B837389F8F5C2 CRC64;

Query Match 13.3%; Score 242.5; DB 2; Length 801;
Best Local Similarity 34.9%; Pred. No. 1.8e-10;
Matches 58; Conservative 21; Mismatches 74; Indels 13; Gaps 6;

QY 30 FSGAELTYLLEVOYQKFEARKMAAGTVSMFLYONGSEINDGRPWVEVDIEVGKNP 89
Db 633 YSGGEFTNNFHYHYGYECGMOAMKNDGVSSFTYTGPS---DNDPMDIEDIEIGKNT 689
QY 90 GSFQSNITTCGAKAQKTSKHNHVAVSPAADQAFHTYGLFEMTPNVRWTVDSQEVKTEGGQ 149
Db 690 TQGFNYFNGVGG--NHHYHGLGDADDFVTYAFEMRPESIRNVRVNGELVHTA---- 742
QY 150 VSNULTGT-OGLRFNLMSSESA-AMVGOFDESKLPLOFIMWVYKYP 193
Db 743 TQDIPKTPGKIMMMLPGVGVDEMTGVRNEDPVTQYDMV---KYTP 787

RESULT 4
Q45691 PRELIMINARY; PRT; 239 AA.
AC 045691;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE ENDO-BETA-1,3-1,4-GLUCANASE.
GN BEG1.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 8565;
RA van Rensburg P., Van Zyl W.H., Pretorius I.S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U60830; AAB05759.1; -
DR HSSP: P27051; IGBG.
DR InterPro: IPR000757; -
DR pfam: PF00722; Glyco_hydro_16; 1.
DR PRINTS: PR00737; GLHYDRLASE16.
DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
SQ SEQUENCE 239 AA; 26969 MW; C4219760D13F878A CRC64;

Query Match 13.1%; Score 238.5; DB 2; Length 239;

Query Match	Best Local Similarity	12.7%; Score 212.5; DB 2; Length 212;
Matches 62; Conservative	26; Mismatches 73; Indels 17; Gaps	
QY 20	ATTNVSAKDPSGAEALTYLLEEVYQKFEARMKMAASGTSSMFLYQNGSELADGRPWVE	79
DB 48	SLTSSAYNNK-FDGGERYSKNTYRYGLYEVRMKPAKNTGIVSFFTYTGP--ANGTQMD	103
QY 80	VDIEVLGNKPGSFQSNITTKRAGAKTSEKHHAVSPADQAHITGILEKTPYVYKRYVDG	139
DB 104	IDIEFLGKDYTKVQPNYYTNGIGH--EKVVDLDFDSSGHHYAFADQPGYIRWYVDG	160
QY 140	QEVRTTEGGVSNLGTQGLRENTLWSSESA-AWVGQFDESKLPLEQFINWKKVYKT	195
DB 161	VLKHT---ATTNPKTFQIIMNLMNGTGVDSWLGPIYGNV-PLVAYEDWV---KYY	210
RESULT 5		
Q9FDC9	PRELIMINARY; PRT; 212 AA.	
AC Q9FDC9	01-MAR-2001 (TREMBLrel. 16, Created)	
DT 01-MAR-2001	(TREMBLrel. 16, Last sequence update)	
DT 01-MAR-2001	(TREMBLrel. 16, Last annotation update)	
DE ENO-1.3-1.4-BETA-GLUCANASE (PRAGMEN)		
OS Paenibacillus polymyxa (Bacillus polymyxa)		
OC Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC Bacillus/Staphylococcus group; Paenibacillus.		
OX NCBI_TaxID=1406;		
RP SEQUENCE FROM N.A.		
RA Yao W., Wang Y., Song W., Yang K., Su Z.;		
RT "Gene cloning of an antifungal protein."		
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.		
DR EMBL; AF284449; AAC02415.1; -		
FT NON TER	1	
FT NON TER	212	
FT SEQUENCE	212 AA; 24127 MW; 7E707E7FEEBF7A440 CRC64;	
Query Match	12.7%; Score 212.5; DB 2; Length 212;	
Best Local Similarity	34.8%; Pred. No. 1.9e-10;	
Matches 62; Conservative	26; Mismatches 73; Indels 17; Gaps	
QY 20	ATTNVSAKDPSGAEALTYLLEEVYQKFEARMKMAASGTSSMFLYQNGSELADGRPWVE	79
DB 48	SLTSSAYNNK-FDGGERYSKNTYRYGLYEVRMKPAKNTGIVSFFTYTGP--ANGTQMD	103
QY 80	VDIEVLGNKPGSFQSNITTKRAGAKTSEKHHAVSPADQAHITGILEKTPYVYKRYVDG	139
DB 104	IDIEFLGKDYTKVQPNYYTNGIGH--EKVVDLDFDSSGHHYAFADQPGYIRWYVDG	160
QY 140	QEVRTTEGGVSNLGTQGLRENTLWSSESA-AWVGQFDESKLPLEQFINWKKVYKT	195
DB 161	VLKHT---ATTNPKTFQIIMNLMNGTGVDSWLGPIYGNV-PLVAYEDWV---KYY	210
RESULT 6		
ID 007856	PRELIMINARY; PRT; 237 AA.	
AC 007856	01-JUL-1997 (TREMBLrel. 04, Created)	
DT 01-JUL-1997	(TREMBLrel. 04, Last sequence update)	
DT 01-MAY-2000	(TREMBLrel. 13, Last annotation update)	
DE BETA-1.3-1.4-GLUCANASE PRECURSOR.		
OS Streptococcus bovis.		
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC Streptococcus.		
OX NCBI_TaxID=1315;		
RP SEQUENCE FROM N.A.		
RA Ekunji M., Flint H.J.;		
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.		

DR	EMBL:	Z92911;	CAB07443.1;	-.
DR	HSSP:	P23904;	IABK.	
DR	InterPro:	IPR000757;	-	
DR	Pfam:	PF00722;	Glyco_hydro_16; 1.	
DR	PRINTS:	PR00737;	GLHYDRLASE16.	
DR	PROSITE:	PS01034;	GLYCOSYL_HYDROL_F16; 1.	
KW	SIGNAL.			
FT	SIGNAL	1	24	POTENTIAL.
FT	CHAIN	25	237	BETA-(1,3-1,4)-GLUCANASE.
SQ	SEQUENCE	237 AA;	26989 MW;	7DEFBCE53790470 CRC64;
<hr/>				
Query Match		12.6%;	Score 230;	DB 2;
Best Local Similarity		33.1%;	Pred. No. 3.5e-10;	
Matches 54;		Conservative 26;	Mismatches 69;	Indels 14;
			Gaps	6;
<hr/>				
QY	30	FSGAELYYLLEVOYGKEFPARKMAASGVSMFLYONGSEIADGRPWVEDIVLGKRP	89	
Dd	81	YTGGWRMSKERFGVGLFVNKKPIKNPCGVSSFYTGPS--DGTKDDEIDIEPLGDT	137	
QY	90	GSPQSNIITGKAQAOKTSEKHHAASPADQAHTYTGLEMTPTVYAWYVDGGEVRTEGGQ	149	
Dd	138	TKVGFNYTT---SGGNHELYLNLFEDDSQGEPHNYGFQMADHTIYWDGRAVYTA----	190	
QY	150	VSNLTFTQG-LRFNLN--SSESAAMVGQFDESKLP.LFOFINMV	189	
Dd	191	YNINPSTPGKITMMNMAMPCTHEVDSMLGAYN-GRTPLVAYDMV	232	
<hr/>				
RESULT	7			
ID	014412	PRELIMINARY;	PRT;	245 AA.
AC	014412			
DT	01-JAN-1998	(TREMBLrel. 05. Created)		
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)		
DT	01-MAY-2000	(TREMBLrel. 13, Last annotation update)		
DE	BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENZO-BETA-1,3-1,4 GLUCANASE) (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHERASE).			
GN	LICA.			
OS	Ophiomyces sp. (Strain PC-2).			
OC	Eukaryota; Fungi; Chytridiomycota; Neocallimastcales;			
OC	Neocallimasticeae; Ophiomyces.			
OX	NCBI_TaxID=37164;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 30-41.			
RX	MEDLINE=97464427; PubMed=9324248;			
RA	Chen H., Li X.-L., Ljungdahl L.G.;			
RT	"Sequencing of a 1,3-1,4-beta-D-glucanase (lichenase) from the anaerobic fungus Ophiomyces strain PC-2: properties of the enzyme expressed in Escherichia coli and evidence that the gene has a bacterial origin."			
RL	J. Bacteriol. 179:6028-6034(1997).			
CC	-1- FUNCTION: HYDROLIZES B-GLUCANS CONTAINING MIXED BETA-1,3 AND BETA-1,4 LINKAGES. THIS ENZYME IS THERMOSTABLE; OPTIMAL TEMPERATURE IS 45 DEGREES CELSIUS AND THE OPTIMUM PH IS 5.8-6.2.			
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.			
DR	EMBL: U63813; RAD04192.1; -.			
DR	HSSP: P23904; IABK.			
DR	InterPro: IPR000757;			
DR	Pfam: PF00722; Glyco_hydro_16; 1.			
DR	PRINTS: PR00737; GLHYDRLASE16.			
DR	PROSITE: PS01034; GLYCOSYL_HYDROL_F16; FALSE_NEG.			
KW	Hydrolase; Glycosidase; Signal.			
FT	SIGNAL	1	29	
FT	CHAIN	30	245	BETA-GLUCANASE.
FT	ACT_SITE	134	134	NUCLEOPHILE (BY SIMILARITY).
FT	ACT_SITE	138	138	PROTON DONOR (BY SIMILARITY).
FT	DISULFID	63	90	BY SIMILARITY.
FT	CONFLICT	36	36	S->L (IN AA SEQUENCE).
SQ	SEQUENCE	245 AA;	27929 MW;	08B5DF57D89F2DDC CXC64;

Query Match 10.4%; Score 190.5; DB 3; Length 245;
 Best Local Similarity 31.6%; Pred. No. 3.8e-07;
 Matches 54; Conservative 17; Mismatches 87; Indels 13; Gaps 6;

QY 21 LTTNVSADKDFSGALTYLEEVQYKFEARKMAAASGVSSMFLYQNGSEIADGRPWVEV 80
 DB 79 LTTDRDGGSTGCGYKRRKNYYGFMVNMKPIKNGVSSFFTYTGPS--DGTKDEI 135
 QY 81 DIEVLGNKPGSFQSNITIGKAGAKTSEKHNHVASPADQAHTYGLTEMTPNYVRYWTDGO 140
 DB 136 DIEFLGVDYTKYQENYNTNGGHH---EHIIYLGFDASQGHYTGFRMARSTIWTYDGT 192
 QY 141 EVKRTGEGQVSNLTGTCG-LRENIMSSESA-AWVGQFDESKLPLFOFINWV 189
 DB 193 AVYTA---YDNIPDTPGKIMMAMNGIGVMDLRFN-GRTNISAYYDMV 238

RESULT 8

Q9K7X6 PRELIMINARY; PRT; 302 AA.

ID Q9K7X6
 AC Q9K7X6
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE HYBRID-EMDO-BETA-1.3-1.4 GLUCANASE.
 GN BH3231.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=8665;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-C-125 / JCM 9153;
 RA Takami H., Nakase K., Takai Y.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF001318; BAB06950.1; -
 DR InterPro: IPR000757; -
 DR Pfam: PF00722; Glyco_hydro.16; 1.
 DR PRINTS: PR00737; GLHYRLASE16.
 DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; UNKNOWN_1.
 SO SEQUENCE 302 AA; 33975 MW; 2EDDA630FD5355CA CRC64;

Query Match 10.4%; Score 190.5; DB 2; Length 302;
 Best Local Similarity 27.8%; Pred. No. 5e-07;
 Matches 62; Conservative 34; Mismatches 98; Indels 29; Gaps 9;

QY 29 DESGAEIYLEEVQYKFEARKMAAASGVSSMFLYQNGSEIADGRPWVEVDIEVLGN 88
 DB 16 NYCEAEYSTNHFGYGLYEVSMKPSNVSGVSSFFTYTGPS--YNGAPMEIDIEFLGND 73
 QY 89 PGFOSNIITGKAGAKTSEKHNHVASPADQAHTYGLTEMTPNYVRYWTDGOEVRKTEGG 148
 DB 74 TTKVQRFYTYNGSG---NEIILYDLGFDANSENRTYAFDMQEHINNYVNGELVATATEN 130
 QY 149 QVSNLTGTCG-LRENIMSSESA-AWVGQFDESKLPLFOFINWVYVYTP---GQG 109
 DB 131 IPSN---PSKMMNINMTYCIDEMAGAYWQNNAS-----YEW---RYTPNNENSQT 178
 QY 200 EGSDFTLMDTDFDPSGRW--GKDMTFDGRVDLDTKNI 240
 DB 179 PLASDFOLHACEYSDARGVSMQGVGSF-YPGSWIKFDVNL 220

RESULT 9

Q9FI31 PRELIMINARY; PRT; 282 AA.

ID Q9FI31
 AC Q9FI31
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE XYLOGLUCAN ENDO-1,4-BETA-D-GLUCANASE.

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;

RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE=9937451; PubMed=10470850;
 RA Kaneo T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
 RA Miyajima N., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
 RT Sequence features of the regions of 1,011,550 bp covered by seventeen
 RT P1 and TAC clones."
 RL DNA Res. 6:183-195(1999).
 DR EMBL: AB017064; BAB11071.1; -
 SO SEQUENCE 282 AA; 32437 MW; FBB8C587F6C9A3C3 CRC64;

Query Match 10.1%; Score 185; DB 10; Length 282;
 Best Local Similarity 27.4%; Pred. No. 1.2e-06;
 Matches 59; Conservative 34; Mismatches 98; Indels 24; Gaps 9;

QY 24 NVSAKDFSGAEIYLEEVQYKFEARKMAA--ASGVSSMFLYQNGSEIADGRPWVEVD 81
 DB 53 SLSDKFGSGGFSHQEFLYKVEVMKLVPGNSAGTYFTYFK-----SPGTYMDEID 106
 QY 82 IEVLGK---NPGSFQSNITIGKAGAKTSEKHNHVASPADQAHTYGLTEMTPNYVRYWTD 138
 DB 107 FEEFGNIGSHPTLHTNWVY-KGTGDEKQFHLMEFDTV--PHYCTIINPQVAVFTID 163
 QY 139 GOEVRKTEGQVSNLT--TGQGLRF--NLWSSESA--WVG---OFDESKLPLFOFINWV 191
 DB 164 GIPREFKNEALGVPPKQPMRLVSLWEAEHWATRGLEMTDMSKAFPTAFYRNWV 223
 QY 192 YKTPGQEGSGDFTLDWTFDPSGRW---RWCK 223
 DB 224 DACVWSNGKSSCSANSSWFTQVLDLFGKKNRVKMAQ 258

RESULT 10

Q42800 PRELIMINARY; PRT; 302 AA.

ID Q42800
 AC Q42800
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE RASP F 9 (FRAGMENT).
 GN RASP F 9.
 OS Aspergillus fumigatus (Sartorya fumigata).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5085;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 42202;
 RA Hermann S.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 42202;
 RX MEDLINE=98141802; PubMed=9482698;
 RA Cramer R.;
 RT "Recombinant Aspergillus fumigatus allergens: from the nucleotide
 RT sequences to clinical applications."
 RL Int. Arch. Allergy Immunol. 115:99-114(1998).
 DR EMBL: AJ23327; CA11266.1; -
 DR HSSP: P23904; IAJK.
 DR InterPro: IPR000757; -
 DR Pfam: PF00722; Glyco_hydro.16; 1.
 KW Allergen.
 FT NON_TER 1 1
 FT NON_TER 302 302


```

Db 323 NIEVGPPTCGEIDIMELHDTRTVYGTAGHPGYSAGSAGVAYHLPEGVDPFSEDFHI 382
QY 124 YGLEMTPNVRYMTVDGOEVRKTEGGVSNLTGQGLRFLNLMSSSE-----SAAW 172
Db 383 FSIEMENDEVEWYVDQGLYHLSKDELAEL-----GLE---WVFDHPPELLINAVGGYMP 435
QY 173 GOFDE-SKLPLEFOFINWVVKYK-----YTPGEGGSDFTLMDTDF 216
Db 436 GYDDETTQFQRMVYIDYIRYKDMNPETITGEVDDCEYEQAOAGPEVYEQINN-GTF 494
QY 217 DGS-----RWKGDWTFDGNRVDLTDKNYSDGMILALTRKGOSEFNGQ 262
Db 495 DEPIVNDQANPDPEWFIWQAGDYGISGARV--SDYGV--TDGYAYTITDPTDTHIQ 549

```

RESULT 13

Q9MOD1 PRELIMINARY: PRT: 277 AA.

```

AC 09MOD1:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE XYLOGLUCAN ENDO-1, 4-BETA-D-GLUCANASE-LIKE PROTEIN.
GN AT4G30290.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxId=3702;
RN (1)
RP SEQUENCE FROM N.A.
RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL161576; CAB81022.1;
DR InterPro: IPR000757;
DR Pfam: PF00722; Glyco_hydro_16; 1.
DR PRINTS: PR00737; GLHYDRLASE16.
DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; UNKNOWN.1.
SQ SEQUENCE 277 AA; 31563 MW; B801F67E0CD72E53 CRC64;

```

Query Match 9.5%; Score 172.5; DB 10; Length 277;
 Best Local Similarity 26.6%; Pred. No. 1.1e-05;
 Matches 54; Conservative 34; Mismatches 74; Indels 41; Gaps 9;

```

QY 24 NVSAKDFSGAELTYLEEVYGRFEARKKMAA--ASGVSSMFLYONGSEIADGRPVVEVD 81
Db 48 SLSLDKSSSGSGROSEFLGKAEVOMKLPVCSAGTVITFLK-----SPETTDEID 101
QY 82 IEVLGR--NPGSFQSNITTKAGAKTSEKHNHVPADQAFHTYGLTEPNYVRYMTVD 138
Db 102 FEELGNISGHPTLHTNVT-KSGDGKEQFHLMPPTAN--FHTYCTIMNPQRIIFTVD 158
QY 139 GOEVRKTEGGVSNLTGQGLR-----NLMSSSAANVG--QFDESKLDFLOF 185
Db 159 GIPIREFMNAE-----SRGVPEPTKQPMRLYASLWEAHEMATRGLEKTDMSKAFETAY 212
QY 186 IN-----WVKVYKTPGGE 200
Db 213 YRNYNEGCVWNGKSVCPANSQ 235

```

RESULT 14
 Q9MOD3 PRELIMINARY: PRT: 646 AA.
 AC 060039:
 DT 01-NOV-1996 (Tremblrel. 01, Created)

```

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE LAMINARINASE (EC 3.2.1.39) (GLUCAN ENDO-1,3-BETA-D-GLUCOSIDASE) ((1-
DE >3)-BETA-GLUCAN ENDONHROLASE) (ENDO-1,3-BETA-GLUCANASE).
GN LICA.
OS Thermotoga neapolitana.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxId=2337;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=42706-MC24;
RA MEDLINE=93356813; PubMed=8352795;
RA Dakhova O., Kurepina N., Zverlov V., Svetlichny V.,
RA Vellikdovorskaya G.;
RT "Cloning and expression in Escherichia coli of Thermotoga neapolitana
RT genes coding for enzymes of carbohydrate substrate degradation.";
RL Biochem. Biophys. Res. Commun. 194:1359-1364(1993).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=42706-MC24;
RX MEDLINE=97312008; PubMed=9168619;
RA Zverlov V.V., Volkov I.Y., Vellikdovorskaya T.V., Schwarz W.H.;
RT Highly thermostable endo-1,3-beta-glucanase (laminarinase) lama from
RT Thermotoga neapolitana: nucleotide sequence of the gene and
RT characterization of the recombinant gene product.";
RL Microbiology 143:1701-1708(1997).
CC -! CATALYTIC ACTIVITY: HYDROLYSIS OF 1,3-BETA-D-GLUCOSIDIC LINKAGES
IN 1,3-BETA-D-GLUCANS.
DR EMBL: Z47974; CAAB8008.1;
DR HSSP: P14090; IULO.
DR InterPro: IPR000757;
DR InterPro: IPR003305;
DR Pfam: PF00722; Glyco_hydro_16; 1.
DR Pfam: PF02018; Cbf_6_2.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 646 AA; 73332 MW; 7FE335509284D75 CRC64;

```

Query Match 9.4%; Score 171; DB 2; Length 646;
 Best Local Similarity 23.4%; Pred. No. 4.1e-05;
 Matches 70; Conservative 50; Mismatches 117; Indels 62; Gaps 16;

```

QY 12 LAAAAAALLTVNSAKDFSGAELTY--LEEVYGRFEARKKMAAASGVSSMFLYONGS 69
Db 269 LVTEARKEQSDSEYGYDYTSARITTEGKFEIKYGIETIRAKLPKKGIMPALMWL--GN 326
QY 70 ELND-GRPV--EVDI-EVLGKNPS-FQSNITTKAGAKTSEKHNHVPAD--QAFHT 123
Db 327 NIEVGPPTCGEIDIMELHDTRTVYGTAGHPGYSAGSAGVAYHLPEGVDPFSEDFHV 386
QY 124 YGLEMTPNVRYMTVDGOEVRKTEGGVSNLTGQGLRFLNLMSSSE-----SAAW 172
Db 387 FSIEMENDEVEWYVDQGLYHLSKDELAEL-----GLE---WVFDHPPELLINAVGGYMP 439
QY 173 GOFDE-SKLPLEFOFINWVVKYK-----YTPGEGGSDFTLMDTDF 216
Db 440 GYDDETTQFQRMVYIDYIRYKDMNPETITGEVDDCEYEQAOAGPEVYEQINN-GTF 498
QY 217 DGS-----RWKGDWTFDGNRVDLTDKNYSDGMILALTRKGOSEFNGQ 262
Db 499 DEPIVNDQANPDPEWFIWQAGDYGISGARV--SDYGV--TDGYAYTITDPTDTHIQ 553

```

RESULT 15
 Q9MOD2 PRELIMINARY: PRT: 282 AA.
 AC 09MOD2:
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE XYLOGLUCAN ENDO-1, 4-BETA-D-GLUCANASE-LIKE PROTEIN.
 GN AT4G30280.
 OS Arabidopsis thaliana (Mouse-ear cress).

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:17:23 ; Search time 45.21 Seconds

(without alignments)
467.989 Million cell updates/sec

Title: US-09-654-652a-3

Perfect score: 349

Sequence: 1 MNIKRTAVKSLAVAAAAA.....AKGAKVNPNGHKRYVNFEEH 349

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 412676 seqs, 60623988 residues

Word size: 0

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :
1: /SID58/gcgdata/geneseq/AA1980.DAT:*
2: /SID58/gcgdata/geneseq/AA1981.DAT:*
3: /SID58/gcgdata/geneseq/AA1982.DAT:*
4: /SID58/gcgdata/geneseq/AA1983.DAT:*
5: /SID58/gcgdata/geneseq/AA1984.DAT:*
6: /SID58/gcgdata/geneseq/AA1985.DAT:*
7: /SID58/gcgdata/geneseq/AA1986.DAT:*
8: /SID58/gcgdata/geneseq/AA1987.DAT:*
9: /SID58/gcgdata/geneseq/AA1988.DAT:*
10: /SID58/gcgdata/geneseq/AA1989.DAT:*
11: /SID58/gcgdata/geneseq/AA1990.DAT:*
12: /SID58/gcgdata/geneseq/AA1991.DAT:*
13: /SID58/gcgdata/geneseq/AA1992.DAT:*
14: /SID58/gcgdata/geneseq/AA1993.DAT:*
15: /SID58/gcgdata/geneseq/AA1994.DAT:*
16: /SID58/gcgdata/geneseq/AA1995.DAT:*
17: /SID58/gcgdata/geneseq/AA1996.DAT:*
18: /SID58/gcgdata/geneseq/AA1997.DAT:*
19: /SID58/gcgdata/geneseq/AA1998.DAT:*
20: /SID58/gcgdata/geneseq/AA1999.DAT:*
21: /SID58/gcgdata/geneseq/AA2000.DAT:*
22: /SID58/gcgdata/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	2.9	144	21	Arabidopsis thaliana
2	10	2.9	209	21	Arabidopsis thaliana
3	10	2.9	268	21	Arabidopsis thaliana
4	10	2.9	319	21	Arabidopsis thaliana
5	10	2.9	322	21	Arabidopsis thaliana
6	10	2.9	333	21	Arabidopsis thaliana
7	10	2.9	384	21	Arabidopsis thaliana
8	10	2.9	387	21	Arabidopsis thaliana
9	9	2.6	337	22	Hairy protein amin
10	8	2.3	33	18	Arctic fish antifr
11	8	2.3	37	13	Antifreeze protein

12	8	2.3	37	20	AAV23879
13	8	2.3	37	20	AAV23880
14	8	2.3	37	20	AAW86157
15	8	2.3	37	20	AAW86158
16	8	2.3	37	20	AAW86156
17	8	2.3	37	21	AAV44713
18	8	2.3	38	12	AAV14375
19	8	2.3	38	20	AAV25427
20	8	2.3	40	11	AAV08077
21	8	2.3	41	11	AAV08080
22	8	2.3	41	11	AAV08082
23	8	2.3	41	11	AAV08086
24	8	2.3	41	11	AAV08097
25	8	2.3	43	20	AAV24058
26	8	2.3	46	11	AAV08098
27	8	2.3	50	17	AAV03591
28	8	2.3	51	11	AAV08078
29	8	2.3	52	11	AAV08081
30	8	2.3	52	11	AAV08083
31	8	2.3	62	11	AAV08079
32	8	2.3	66	11	AAV08067
33	8	2.3	73	11	AAV08068
34	8	2.3	82	20	AAV23877
35	8	2.3	97	20	AAV86160
36	8	2.3	114	20	AAV60122
37	8	2.3	131	20	AAV23876
38	8	2.3	131	20	AAV95195
39	8	2.3	162	21	AAV33452
40	8	2.3	165	22	AAV63414
41	8	2.3	219	22	AAV72164
42	8	2.3	231	22	AAV68902
43	8	2.3	231	22	AAV68903
44	8	2.3	231	22	AAV68904
45	8	2.3	231	22	AAV68905

ALIGNMENTS

RESULT	1	
AAV45466	standard; Protein; 144 AA.	
AC	AAV45466;	
XX		
DT	18-OCT-2000 (first entry)	
XX		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 57085.	
XX		
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX		
OS	Arabidopsis thaliana.	
XX		
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000; 2000EP-0301439.	
XX		
PR	25-FEB-1999; 99US-0121825.	
PR	05-MAR-1999; 99US-0123180.	
PR	09-MAR-1999; 99US-0123548.	
PR	23-MAR-1999; 99US-0125788.	
PR	25-MAR-1999; 99US-0126264.	
PR	29-MAR-1999; 99US-0126785.	
PR	01-APR-1999; 99US-0127462.	
PR	06-APR-1999; 99US-0128234.	
PR	08-APR-1999; 99US-0128714.	
PR	16-APR-1999; 99US-0129845.	
PR	19-APR-1999; 99US-0130077.	
PR	21-APR-1999; 99US-0130449.	

Protein derived fr
Protein derived fr
P. americanus anti
P. americanus anti
P. americanus anti
P. americanus anti
Winter flounder 11
Melting pt. dePRES
P. americanus anti
Synthetic antifree
Synthetic antifree
Synthetic antifree
Synthetic antifree
Synthetic antifree
Human alpha 2 C4 a
Synthetic antifree
Synthetic antifree
Synthetic antifree
Synthetic antifree
Synthetic antifree
Winter flounder an
P. americanus anti
Human endometrium
A hypothetical mat
Hypothetical hybr
Zea mays protein f
Human breast cance
Human RNA metaboli
Neisseria meningit
Neisseria meningit
Neisseria meningit

PR 23-APR-1999; 9905-0130510.
 PR 23-APR-1999; 9905-0130891.
 PR 28-APR-1999; 9905-0131449.
 PR 30-APR-1999; 9905-0132048.
 PR 30-APR-1999; 9905-0132407.
 PR 04-MAY-1999; 9905-0132484.
 PR 05-MAY-1999; 9905-0132485.
 PR 06-MAY-1999; 9905-0132486.
 PR 06-MAY-1999; 9905-0132487.
 PR 07-MAY-1999; 9905-0132483.
 PR 11-MAY-1999; 9905-0134256.
 PR 14-MAY-1999; 9905-0134218.
 PR 14-MAY-1999; 9905-0134219.
 PR 14-MAY-1999; 9905-0134221.
 PR 18-MAY-1999; 9905-0134370.
 PR 19-MAY-1999; 9905-0134768.
 PR 20-MAY-1999; 9905-0134941.
 PR 20-MAY-1999; 9905-0135124.
 PR 21-MAY-1999; 9905-0135353.
 PR 24-MAY-1999; 9905-0135629.
 PR 25-MAY-1999; 9905-0136021.
 PR 27-MAY-1999; 9905-0136392.
 PR 28-MAY-1999; 9905-0136782.
 PR 01-JUN-1999; 9905-0137222.
 PR 03-JUN-1999; 9905-0137528.
 PR 04-JUN-1999; 9905-0137502.
 PR 07-JUN-1999; 9905-0137724.
 PR 08-JUN-1999; 9905-0138094.
 PR 10-JUN-1999; 9905-0138540.
 PR 10-JUN-1999; 9905-0138847.
 PR 14-JUN-1999; 9905-0139119.
 PR 16-JUN-1999; 9905-0139452.
 PR 16-JUN-1999; 9905-0139453.
 PR 17-JUN-1999; 9905-0139453.
 PR 18-JUN-1999; 9905-0139454.
 PR 18-JUN-1999; 9905-0139455.
 PR 18-JUN-1999; 9905-0139456.
 PR 18-JUN-1999; 9905-0139457.
 PR 18-JUN-1999; 9905-0139458.
 PR 18-JUN-1999; 9905-0139459.
 PR 18-JUN-1999; 9905-0139460.
 PR 18-JUN-1999; 9905-0139461.
 PR 18-JUN-1999; 9905-0139462.
 PR 18-JUN-1999; 9905-0139463.
 PR 18-JUN-1999; 9905-0139750.
 PR 18-JUN-1999; 9905-0139763.
 PR 21-JUN-1999; 9905-0139817.
 PR 22-JUN-1999; 9905-0139899.
 PR 23-JUN-1999; 9905-0140353.
 PR 23-JUN-1999; 9905-0140354.
 PR 24-JUN-1999; 9905-0140695.
 PR 28-JUN-1999; 9905-0140823.
 PR 29-JUN-1999; 9905-0140891.
 PR 30-JUN-1999; 9905-0141287.
 PR 01-JUL-1999; 9905-0141842.
 PR 01-JUL-1999; 9905-0142154.
 PR 02-JUL-1999; 9905-0142055.
 PR 06-JUL-1999; 9905-0142390.
 PR 08-JUL-1999; 9905-0142803.
 PR 09-JUL-1999; 9905-0142920.
 PR 12-JUL-1999; 9905-0142927.
 PR 13-JUL-1999; 9905-0143542.
 PR 14-JUL-1999; 9905-0143624.
 PR 15-JUL-1999; 9905-0144005.
 PR 16-JUL-1999; 9905-0144085.
 PR 16-JUL-1999; 9905-0144086.
 PR 19-JUL-1999; 9905-0144325.
 PR 19-JUL-1999; 9905-0144331.
 PR 19-JUL-1999; 9905-0144332.
 PR 19-JUL-1999; 9905-0144333.
 PR 19-JUL-1999; 9905-0144334.
 PR 19-JUL-1999; 9905-0144335.
 PR 20-JUL-1999; 9905-0144352.

PR 20-JUL-1999; 9905-0144632.
 PR 20-JUL-1999; 9905-0144884.
 PR 21-JUL-1999; 9905-0144814.
 PR 21-JUL-1999; 9905-0145086.
 PR 21-JUL-1999; 9905-0145088.
 PR 22-JUL-1999; 9905-0145085.
 PR 22-JUL-1999; 9905-0145087.
 PR 22-JUL-1999; 9905-0145089.
 PR 22-JUL-1999; 9905-0145192.
 PR 23-JUL-1999; 9905-0145145.
 PR 23-JUL-1999; 9905-0145218.
 PR 23-JUL-1999; 9905-0145224.
 PR 26-JUL-1999; 9905-0145276.
 PR 27-JUL-1999; 9905-0145913.
 PR 27-JUL-1999; 9905-0145918.
 PR 27-JUL-1999; 9905-0145919.
 PR 28-JUL-1999; 9905-0145951.
 PR 02-AUG-1999; 9905-0146386.
 PR 02-AUG-1999; 9905-0146388.
 PR 02-AUG-1999; 9905-0146389.
 PR 03-AUG-1999; 9905-0147038.
 PR 04-AUG-1999; 9905-0147204.
 PR 04-AUG-1999; 9905-0147302.
 PR 05-AUG-1999; 9905-0147192.
 PR 05-AUG-1999; 9905-0147260.
 PR 06-AUG-1999; 9905-0147303.
 PR 06-AUG-1999; 9905-0147416.
 PR 09-AUG-1999; 9905-0147493.
 PR 09-AUG-1999; 9905-0147935.
 PR 10-AUG-1999; 9905-0148171.
 PR 11-AUG-1999; 9905-0148319.
 PR 12-AUG-1999; 9905-0148341.
 PR 13-AUG-1999; 9905-0148565.
 PR 13-AUG-1999; 9905-0148684.
 PR 16-AUG-1999; 9905-0149368.
 PR 17-AUG-1999; 9905-0149175.
 PR 18-AUG-1999; 9905-0149426.
 PR 20-AUG-1999; 9905-0149722.
 PR 20-AUG-1999; 9905-0149723.
 PR 20-AUG-1999; 9905-0149929.
 PR 23-AUG-1999; 9905-0149930.
 PR 23-AUG-1999; 9905-0149930.
 PR 25-AUG-1999; 9905-0150566.
 PR 26-AUG-1999; 9905-0150884.
 PR 27-AUG-1999; 9905-0151065.
 PR 27-AUG-1999; 9905-0151066.
 PR 27-AUG-1999; 9905-0151080.
 PR 30-AUG-1999; 9905-0151303.
 PR 31-AUG-1999; 9905-0151438.
 PR 01-SEP-1999; 9905-0151630.
 PR 07-SEP-1999; 9905-0152363.
 PR 10-SEP-1999; 9905-0153070.
 PR 13-SEP-1999; 9905-0153758.
 PR 15-SEP-1999; 9905-0154018.
 PR 16-SEP-1999; 9905-0154039.
 PR 20-SEP-1999; 9905-0154779.
 PR 22-SEP-1999; 9905-0155139.
 PR 22-SEP-1999; 9905-0155486.
 PR 24-SEP-1999; 9905-0156359.
 PR 28-SEP-1999; 9905-0156458.
 PR 29-SEP-1999; 9905-0156596.
 PR 04-OCT-1999; 9905-0157117.
 PR 05-OCT-1999; 9905-0157753.
 PR 06-OCT-1999; 9905-0157865.
 PR 07-OCT-1999; 9905-0158029.
 PR 08-OCT-1999; 9905-0158232.
 PR 12-OCT-1999; 9905-0158369.
 PR 13-OCT-1999; 9905-0159293.
 PR 13-OCT-1999; 9905-0159294.
 PR 13-OCT-1999; 9905-0159295.
 PR 14-OCT-1999; 9905-0159329.
 PR 14-OCT-1999; 9905-0159330.
 PR 14-OCT-1999; 9905-0159331.

PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.98; Score 10; DB 21; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 SALVAAAAA 19
Db 5 salavaaaa 14

RESULT 2

AAAG45465 ID AAG45465 standard; Protein: 209 AA.

AC AAG45465;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 57084.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.

XX Arabidopsis thaliana.

OS EP1033405-A2.

PN 06-SEP-2000.

PD 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126284.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 30-APR-1999; 99US-0132407.

XX 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148564.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150684.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157713.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158365.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161040.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.9%; Score 10; DB 21; Length 209;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 SALAVAAAA 19
Db 70 salavaaaa 79

RESULT 3
AAC49854
ID AAC49854 standard; Protein; 268 AA.
XX AAC49854;
AC
AC
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63112.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
PD
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.

PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.9%; Score 10; DB 21; Length 268;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 SALAVAAAAA 19
| | | | | | | | | |
Db 5 salavaaaaa 14

RESULT 4
AAG19454
ID AAG19454 standard; Protein; 319 AA.
XX
AC AAG19454;
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 21262.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
PN EP1033405-A2.
PD
XX 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154179.
PR 20-SEP-1999; 99US-0155119.
PR 22-SEP-1999; 99US-0155486.
PR 23-SEP-1999; 99US-0155659.
PR 24-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0157117.
PR 04-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157865.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.

PR 29-OCT-1999; 99US-0162142.
Query Match 2.9%; Score 10; DB 21; Length 319;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 10 SALAVAAAAA 19
|||||
Db 5 salavaaaa 14
RESULT 5
AAG49849
ID AAG49849 standard; Protein; 322 AA.
XX
AC AAG49849;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63105.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126786.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139482.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145961.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151085.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151087.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.9%; Score 10; DB 21; Length 322;
Best Local Similarity 100.0%; Fred. No. 0.31; 0; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 10 SALAVAAAA 19

DB 5 salavaaaa 14
|||||||
RESULT 6
AAG49853
ID AAG49853 standard; Protein: 333 AA.
XX AAG49853;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63111.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
FE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130549.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145115.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147410.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

```
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149910.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158222.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159684.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161892.
PR 28-OCT-1999; 99US-0161893.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 2.9%; Score 10; DB 21; Length 333;

Best Local Similarity 100.0%; Pred No. 0.32; Mismatches 0; Indels 0; Gaps 0;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 SALAVAAAAA 19
| | | | | | | | | |
Db 70 salavaaaaaa 79

RESULT 7
AAG19453
ID AAG19453 standard; Protein; 384 AA.

```
XX XX  
AC AAG19453;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 21261.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134321.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.
```

```
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146387.
PR 03-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158363.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.9%; Score 10; DB 21; Length 384;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 SALAVAAAAA 19
DB 70 salavaaaaa 79

RESULT 8
ID AAG49848 standard; Protein; 387 AA.
AC AAG49848;
XX
XX 18-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana protein fragment SEQ ID NO: 63104.
DE
XX
```

KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 9905-0121825.
XX 05-MAR-1999; 9905-0123380.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 28-APR-1999; 9905-0130891.
PR 30-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 04-MAY-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132487.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134321.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.

PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 02-JUL-1999; 9905-0142154.
PR 06-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 21-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145716.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145951.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.

PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 22-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 2.9%; Score 10; DB 21; Length 387;
 Best Local Similarity 100.0%; Pred. No. 0.37;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 SALAVAAAAA 19
 |||||
 DB 70 salavaaaa 79

RESULT 9
 AAB72380
 ID AAB72380 standard; protein: 337 AA.

AC AAB72380;

DT 22-MAY-2001 (first entry)

DE Hairly protein amino acid sequence.

KW Cardiac helix-loop-helix factor; CHF; transcription factor;

KW cardiomyocyte; proliferation; myocarditis; myocardial infarction;

KW cardiomyocyte regeneration; angiogenesis inhibitor; differentiation;

KW smooth muscle cell growth; heart disease; hairy.

XX Drosophila sp.

OS
 XX

PN WO200112126-A2.
 XX
 PD 22-FEB-2001.
 XX
 PF 11-AUG-2000; 2000WO-US21858.
 XX
 PR 13-AUG-1999; 99US-0148974.
 XX
 PA (HARD) HARVARD COLLEGE.
 XX (LEEM/) LEE M.
 PI Lee M, Chin MT;
 XX
 DR WPI: 2001-211119/21.
 XX
 PT Novel cardiovascular helix-loop-helix factor polypeptide and
 PT polynucleotide useful for regenerating heart tissue and promoting
 XX smooth muscle differentiation to treat heart disease or heart injury -
 PS Disclosure: Fig 1; 46pp; English.

This invention relates to human cardiovascular helix-loop-helix factor (CHF) polypeptides. CHF is a transcription factor expressed in cardiovascular tissue, which contains a basic helix-loop-helix domain. CHF-1 shares structural homology with hairy, a Drosophila protein essential for the development of the peripheral nervous system. The CHF polypeptide is useful for expanding an endogenous population of proliferative cardiomyocytes in a heart tissue, promoting proliferation of a cardiomyocyte in a mammal having myocarditis or that which has suffered myocardial infarction and for regenerating cardiomyocytes in vivo. DNA encoding a CHF-1 polypeptide is also useful for promoting proliferation of a cardiomyocyte, regenerating cardiomyocytes in vitro, inhibiting angiogenesis in a tissue, inducing differentiation of smooth muscle cells in a mammalian tissue, preferably venous tissue, inducing growth of smooth muscle cells in a vein explant, promoting smooth muscle cell regeneration in an injured or diseased vascular tissue (venous or arterial) and for reducing vein graft stenosis in a mammal by contacting the tissue ex vivo with the DNA prior to implantation of the tissue into an artery of the mammal. The present sequence represents the hairy protein from Drosophila, with which CHF-1 shares homology.

Sequence 337 AA;

Query Match 2.6%; Score 9; DB 22; Length 337;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 LAVAAAAA 20
 |||||
 DB 238 lavaaaaa 246

RESULT 10
 AAM27490
 ID AAM27490 standard; peptide: 33 AA.

AC AAM27490;

DT 20-APR-1998 (first entry)

DE Arctic fish antifreeze peptide.

KW Antifreeze protein; cryoprotectant; arctic fish; transfusion;

KW haematopoietic cell; freeze-drying.

XX Pisces.

OS DE19611969-A1.

XX 02-OCT-1997.

PD 26-MAR-1996; 96DE-1011969.

XX 26-MAR-1996; 96DE-1011969.
 XX (RADU/) RADULESCU R T.
 XX
 XX Radulescu RT;
 XX
 XX WPI; 1997-481546/45.
 XX
 XX Use of antifreeze peptide in transfusion medicine - especially as
 PT cryoprotectant for haematopoietic cells
 XX
 XX Disclosure; Column 1; 2pp; German.
 XX
 XX This sequence represents an antifreeze peptide which can be used in the
 CC field of transfusion medicine. This peptide can be used as a
 CC cryoprotectant to protect haematopoietic cells during freezing, thawing
 CC and freeze-drying. The peptide has the same activity as the native
 CC arctic fish protein from which it is derived and should have no
 CC significant side effects as it rapidly degrades in vivo.
 XX
 SO Sequence 33 AA;

Query Match 2.3%; Score 8; DB 18; Length 33;
 Best Local Similarity 100.0%; Pred. No. 2.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 AAAAALT 22
 |||||
 DB 2 aaaaalt 9

RESULT 11

AA26105
 ID AAR26105 standard; Protein; 37 AA.
 XX
 AC AAR26105;
 XX
 DT 01-FEB-1993 (first entry)
 XX
 DE Antifreeze protein from Winter Flounder.
 XX
 KM Thermal hysteresis protein; Ice crystal.
 XX
 OS Pseudopleuronectes americanus.
 XX
 PN WO9212722-A.
 XX
 PD 06-AUG-1992.
 XX
 PF 17-JAN-1992; 92WO-US00452.
 XX
 PR 17-JAN-1991; 91WO-US00351.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Devries AL, Rubinsky B;
 XX
 DR WPI; 1992-284413/34.
 XX
 PT Compns. contg. thermal hysteresis protein - used to protect and
 PT preserve biological materials in non-physiological temps. and
 PT conditions
 XX
 PS Claim 28; Page 21; 126pp; English.
 XX
 CC The antifreeze protein from the Winter Flounder lacks sugars (C.f.
 CC the antifreeze glycopeptides of Nototheniidae fish, e.g. AAR26104)
 CC and instead has high percentages of hydrophilic amino acids (Thr
 CC and Asp) while still retaining a large amount of Ala. The AFP from
 CC P.americanus is one of the preferred thermal hysteresis proteins for
 CC use in the compositions of the invention.

CC See also AAR26106.
 XX
 SO Sequence 37 AA;

Query Match 2.3%; Score 8; DB 13; Length 37;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 AAAAALT 22
 |||||
 DB 6 aaaaalt 13

RESULT 12

AA23879
 ID AAY23879 standard; Protein; 37 AA.
 XX
 AC AAY23879;
 XX
 DT 06-OCT-1999 (first entry)
 XX
 DE Protein derived from winter flounder antifreeze protein (AFP) A.
 XX
 KM Antifreeze protein; AFP; Ice crystal; Winter flounder;
 KM Pseudopleuronectes americanus; Ice crystal growth; frozen product;
 KM ice cream; paint; transplant organ; yeast culture; brewing; baking;
 KM frozen.
 XX
 OS Synthetic.
 OS Pseudopleuronectes americanus.
 XX
 PN US5928877-A.
 XX
 PD 27-JUL-1999.
 XX
 PF 20-NOV-1997; 97US-0975166.
 XX
 PR 20-JUL-1992; 92US-0917216.
 PR 19-SEP-1989; 89US-0409217.
 PR 28-FEB-1990; 90US-0486333.
 PR 12-JAN-1994; 94US-0180524.
 PR 20-NOV-1997; 97US-0975166.
 XX
 PA (MILL-) MILLER BREWING CO.
 XX
 PI Cronan CL, Lusk L;
 XX
 DR WPI; 1999-429496/36.
 XX
 PT Detecting the presence of antifreeze proteins in samples
 XX
 PS Example 1; Column 14; 24pp; English.
 XX

The present sequence is derived from an antifreeze protein (AFP) A. The
 CC specification describes a method for detecting the presence of
 CC functional AFPs in samples. The method comprises freezing a test
 CC solution which contains sucrose, has a melting point (mp), and in
 CC which ice crystals are formed, raising the temperature of the test
 CC solution to its mp to partially melt the ice crystals, lowering the
 CC temperature 1 to 2 degrees Celsius below the mp, which causes the
 CC partially melted crystals to begin to grow, and determining the
 CC presence of functional AFPs by observing the sample for the growth
 CC of hexagonal ice crystals. The method may be used for detecting the
 CC presence of functional AFPs in samples. AFPs (which may be isolated
 CC from fish in cold polar marine waters such as the winter flounder
 CC (Pseudopleuronectes americanus) bind to, and inhibit the formation
 CC of ice crystals in water as it is frozen. They may be used to
 CC prevent ice crystal growth in a wide range of frozen products such
 CC as ice cream and paint. In particular they may be applied to living
 CC cells and tissues (e.g. transplant organs and yeast cultures for
 CC brewing and baking), allowing them to be frozen without the risk of
 CC ice crystal formation, which disrupts cellular membranes and kills

CC the cells making them useless.
 XX
 SQ Sequence 37 AA;

Query Match 2.3%; Score 8; DB 20; Length 37;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 AAAAAALT 22
 |||||
 Db 6 aaaaalt 13

RESULT 13
 AAY23880
 ID AAY23880 standard; Protein; 37 AA.

XX AAY23880;

DT 06-OCT-1999 (first entry)

DE Protein derived from winter flounder antifreeze protein (AFP) A.

XX Antifreeze protein; AFP; ice crystal; winter flounder;

KW pseudopleuronectus americanus; ice crystal growth; frozen product;

KM ice cream; paint; transplant organ; yeast culture; brewing; baking;

KM frozen.

XX Synthetic.

OS Pseudopleuronectus americanus.

XX US592887-A.

PN 27-JUL-1999.

XX 20-NOV-1997; 97US-0975166.

XX 20-JUL-1992; 92US-0917216.

PR 19-SEP-1989; 89US-0409217.

PR 28-FEB-1990; 90US-0486333.

PR 12-JAN-1994; 94US-0180524.

PR 20-NOV-1997; 97US-0975166.

XX (MILL-) MILLER BREWING CO.

PA Cronan CL, Lusk L;

PI WPI; 1999-429496/36.

XX Detecting the presence of antifreeze proteins in samples

XX Example 1; Column 14; 24pp; English.

CC The present sequence is derived from an antifreeze protein (AFP) A. The
 CC specification describes a method for detecting the presence of
 CC functional AFPs in samples. The method comprises freezing a test
 CC solution which contains sucrose, has a melting point (mp), and in
 CC which ice crystals are formed, raising the temperature of the test
 CC solution to its mp to partially melt the ice crystals, lowering the
 CC temperature 1 to 2 degrees Celsius below the mp, which causes the
 CC partially melted crystals to begin to grow, and determining the
 CC presence of functional AFPs by observing the sample for the growth
 CC of hexagonal ice crystals. The method may be used for detecting the
 CC presence of functional AFPs in samples. AFPs (which may be isolated
 CC from fish in cold polar marine waters such as the winter flounder
 CC (Pseudopleuronectus americanus) bind to, and inhibit the formation
 CC of ice crystals in water as it is frozen. They may be used to
 CC prevent ice crystal growth in a wide range of frozen products such
 CC as ice cream and paint. In particular they may be applied to living
 CC cells and tissues (e.g. transplant organs and yeast cultures for
 CC brewing and baking), allowing them to be frozen without the risk of
 CC ice crystal formation, which disrupts cellular membranes and kills

CC the cells making them useless.
 XX
 SQ Sequence 37 AA;

Query Match 2.3%; Score 8; DB 20; Length 37;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 AAAAAALT 22
 |||||
 Db 6 aaaaalt 13

RESULT 14
 AAW86157
 ID AAW86157 standard; protein; 37 AA.

XX AAW86157;

DT 11-MAR-1999 (first entry)

DE P. americanus antifreeze protein A variant AFP A(10).

XX Winter flounder; AFP A; antifreeze protein; transformation; yeast;

KW Saccharomyces cerevisiae; variant; AFP; ice crystal; food; beverage.

XX Pseudopleuronectus americanus.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 23

FT /label=L23D

XX /note="wild-type Leu is replaced by Asp"

XX US5849537-A.

XX 15-DEC-1998.

XX 12-JAN-1994; 94US-0180524.

XX 20-JUL-1992; 92US-0917216.

PR 19-SEP-1989; 89US-0409217.

PR 28-FEB-1990; 90US-0486333.

PR 12-JAN-1994; 94US-0180524.

XX (MILL-) MILLER BREWING CO.

PA Barney MC, Bower PA, Chicoye E, Cronan CL, Hulge N;

PI Kot E, Lusk L, Rhodes T, Tripp M;

XX WPI; 1999-094410/08.

XX Vector for expressing antifreeze protein in yeast - for producing

XX variant of Pseudopleuronectus americanus antifreeze protein

XX Claim 2; Column 14; 25pp; English.

XX The present sequence represents a variant of the antifreeze protein A

XX (AFP A). The invention relates to an expression vector for transforming

XX yeast (Saccharomyces cerevisiae) so that it expresses and secretes an

XX antifreeze protein with no N-terminal methionine. The vector contains a

XX DNA sequence encoding a variant of the winter flounder AFP A

XX (Pseudopleuronectus americanus antifreeze protein A) sequence in which

XX one amino acid is replaced with a different amino acid. The antifreeze

XX protein can be used to reduce the rate of ice crystal growth in food,

XX beverage and nonfood applications.

SQ Sequence 37 AA;

Query Match 2.3%; Score 8; DB 20; Length 37;
 Best Local Similarity 100.0%; Pred. No. 3.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 15 AAAAALT 22
 |||||
 Db 6 aaaaalt 13

Search completed: August 23, 2001, 11:17:24
 Job time: 196 sec

RESULT 15
 AAW86158
 ID AAW86158 standard; protein; 37 AA.
 XX
 AC AAW86158;
 XX
 DT 11-MAR-1999 (first entry)
 XX
 DE P. americanus antifreeze protein A variant AFP A(38).
 XX
 KM Winter flounder; AFP A; antifreeze protein; transformation; yeast;
 KM Saccharomyces cerevisiae; variant; AFP; ice crystal; food; beverage.
 XX
 OS Pseudopleuronectes americanus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 36 /label= A36V
 FT /note= "wild-type Ala is replaced by Val"
 XX
 PN US5849537-A.
 XX
 PD 15-DEC-1998.
 XX
 PF 12-JAN-1994; 94US-0180524.
 XX
 PR 20-JUL-1992; 92US-0917216.
 PR 19-SEP-1989; 89US-0409217.
 PR 28-FEB-1990; 90US-0486333.
 PR 12-JAN-1994; 94US-0180524.
 XX
 PA (MILL-) MILLER BREWING CO.
 XX
 PI Barney MC, Bower PA, Chitoye E, Cronan CL, Huige N;
 PI Kot E, Lusk L, Rhodes T, Tripp M;
 XX
 DR MPI; 1999-094410/08.
 DR N-PSDB; AAV84680.
 XX
 PT Vector for expressing antifreeze protein in yeast - for producing
 PT variant of Pseudopleuronectes americanus antifreeze protein
 XX
 PS Example 1; Column 14; 25pp; English.
 XX
 CC The present sequence represents a variant of the antifreeze protein A
 CC (AFP A). The invention relates to an expression vector for transforming
 CC yeast (Saccharomyces cerevisiae) so that it expresses and secretes an
 CC antifreeze protein with no N-terminal methionine. The vector contains a
 CC DNA sequence encoding a variant of the winter flounder AFP A
 CC (Pseudopleuronectes americanus antifreeze protein A) sequence in which
 CC one amino acid is replaced with a different amino acid. The antifreeze
 CC protein can be used to reduce the rate of ice crystal growth in food,
 CC beverage and nonfood applications.
 CC
 SQ Sequence 37 AA;

Query Match 2.3%; Score 8; DB 20; Length 37;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 AAAAALT 22
 |||||
 Db 6 aaaaalt 13

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:18:03 ; Search time 32.82 Seconds
(without alignments)
810.022 Million cell updates/sec

Title: US-09-654-652a-3
Perfect score: 349
Sequence: 1 MNIKKTAVKASALAVAAAAA.....AKGAKVNPNGHKRYRVNFEEH 349

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	349	100.0	349	2 A44507	licheninase (EC 3.1.1.1)
2	10	2.9	348	2 T52635	mitogen-activated
3	10	2.9	348	2 T51340	mitogen-activated
4	10	2.9	621	1 S59632	endo-1,4-beta-xyla
5	9	2.6	184	2 B84259	hypothetical prote
6	9	2.6	337	2 S06956	segmentation prote
7	9	2.6	439	2 A72599	hypothetical prote
8	9	2.6	537	1 A46600	methylnalonalde-sem
9	9	2.6	776	2 T29064	hyaluronate lyase
10	9	2.6	918	2 T48719	protein kinase C (
11	9	2.6	1475	1 S42718	nuclear pore compl
12	8	2.3	37	1 FDFL3W	antifreeze protein
13	8	2.3	45	2 PM0589	tyrosine 3-monooxy
14	8	2.3	45	2 PM0590	tyrosine 3-monooxy
15	8	2.3	45	2 PM0591	tyrosine 3-monooxy
16	8	2.3	45	2 PM0592	tyrosine 3-monooxy
17	8	2.3	45	2 PM0593	tyrosine 3-monooxy
18	8	2.3	61	2 G41476	antifreeze antigen 7
19	8	2.3	82	1 FDFLAW	antifreeze protein
20	8	2.3	82	2 S02326	antifreeze protein
21	8	2.3	82	2 JS0706	antifreeze protein
22	8	2.3	82	2 A05161	antifreeze protein
23	8	2.3	82	2 JS0705	antifreeze protein
24	8	2.3	82	2 JS1125	antifreeze protein
25	8	2.3	158	2 T34030	hypothetical prote
26	8	2.3	219	2 T51382	achaeate-scute homo
27	8	2.3	231	2 E81215	thiol-disulfide in
28	8	2.3	284	2 T06141	probable receptor-
29	8	2.3	291	2 T51668	myb-related transc

30	8	2.3	317	2 C70874	hypothetical prote
31	8	2.3	339	2 T26328	hypothetical prote
32	8	2.3	368	2 G83463	probable methyltra
33	8	2.3	368	2 T46615	chemotaxis protein
34	8	2.3	374	2 T03875	probable homeobox
35	8	2.3	375	2 T03874	probable homeobox
36	8	2.3	378	2 T06512	DNA-binding protei
37	8	2.3	381	2 S29560	fructose-bisphosph
38	8	2.3	392	2 B48423	homeoic protein e
39	8	2.3	401	2 A48423	engrailed homeod
40	8	2.3	407	2 T06408	probable fructose-
41	8	2.3	414	2 A86229	hypothetical prote
42	8	2.3	424	2 S09884	hypothetical prote
43	8	2.3	429	2 T07815	S-lucos-specific g
44	8	2.3	429	2 D70784	probable Rieske [2
45	8	2.3	459	2 S13064	Id-myo-inositol-tr

ALIGNMENTS

RESULT 1
A44507
licheninase (EC 3.2.1.73) - Fibrobacter succinogenes
C:Species: Fibrobacter succinogenes
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 08-Oct-1999
C:Accession: A44507
R:Reather, R.M., Erfile, J.D.
J. Bacteriol. 172, 3837-3841, 1990
A:Title: DNA sequence of a Fibrobacter succinogenes mixed-linkage beta-glucanase (1.3
A:Reference number: A44507; MUID:90299807
A:Accession: A44507
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <TEA>
A:Cross-references: EMBL:M3676; NID:q148575; PIDN:AAA24896.1; PID:q148576
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match	100.0%	Score 349	DB 2	Length 349
Best Local Similarity	100.0%	Pred. No. 0		
Matches 349	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1 MNIKKTAVKASALAVAAAAAALTTNWSAKDFSGAEYTLTEVOYGRFEARKMAAASGVTS 60			
DB	1 MNIKKTAVKASALAVAAAAAALTTNWSAKDFSGAEYTLTEVOYGRFEARKMAAASGVTS 60			
QY	61 SMFLYQNGSEIADGRPWVEVDIEVLGKNPQSFQSNITGKAGAQRTSEKHNAVSPADQA 120			
DB	61 SMFLYQNGSEIADGRPWVEVDIEVLGKNPQSFQSNITGKAGAQRTSEKHNAVSPADQA 120			
QY	121 FTTTGLEWTPNVTVDGQEVKRTGEGGVSNLTGQGLRFMLNSESAAVGGODESKL 180			
DB	121 FTTTGLEWTPNVTVDGQEVKRTGEGGVSNLTGQGLRFMLNSESAAVGGODESKL 180			
QY	121 FTTTGLEWTPNVTVDGQEVKRTGEGGVSNLTGQGLRFMLNSESAAVGGODESKL 180			
DB	121 FTTTGLEWTPNVTVDGQEVKRTGEGGVSNLTGQGLRFMLNSESAAVGGODESKL 180			
QY	181 PTFQFINWVKYKTPPGQEGGSDFTLDTQNFDFDGSRMKGMTFPGNVDLTDKNI 240			
DB	181 PTFQFINWVKYKTPPGQEGGSDFTLDTQNFDFDGSRMKGMTFPGNVDLTDKNI 240			
QY	181 PTFQFINWVKYKTPPGQEGGSDFTLDTQNFDFDGSRMKGMTFPGNVDLTDKNI 240			
DB	181 PTFQFINWVKYKTPPGQEGGSDFTLDTQNFDFDGSRMKGMTFPGNVDLTDKNI 240			
QY	241 YSRDMLLILALTRKQESFNGVPRDDEPAPOSSSSAPASSSSVPASSSSSAFV 300			
DB	241 YSRDMLLILALTRKQESFNGVPRDDEPAPOSSSSAPASSSSVPASSSSSAFV 300			
QY	241 YSRDMLLILALTRKQESFNGVPRDDEPAPOSSSSAPASSSSVPASSSSSAFV 300			
DB	241 YSRDMLLILALTRKQESFNGVPRDDEPAPOSSSSAPASSSSVPASSSSSAFV 300			
QY	301 PSSSSATNATIGMRTTPAVAKEHNLVNAKAKVNPNGHKRYRVNFEEH 349			
DB	301 PSSSSATNATIGMRTTPAVAKEHNLVNAKAKVNPNGHKRYRVNFEEH 349			
RESULT 2				
T52635	mitogen-activated protein kinase kinase (EC 2.7.1.1) alpha [imported] - Arabidopsis t			
C:Species:	Arabidopsis thaliana (mouse-ear cress)			
C:Date:	24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 03-Nov-2000			

C:Accession: T52635
R:Hamai, A.; Jouanin, S.; Leprince, S.; Kreis, M.; Henry, Y.
Plant Sci. 140, 41-52, 1999
A:Title: Molecular characterization and expression of an Arabidopsis Thaliana L. MAP kin
A:Reference number: 226147
A:Accession: T52635
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-348 <HAM>
A:Cross-references: EMBL:Y07694; PIDN:CAA68958.1
A:Experimental source: cultivar Columbia; seedling
C:Genetics:
A:Gene: MAP2Kalpha
A:Map position: 3
C:Superfamily: kinase-related transforming protein; protein kinase homology
C:Keywords: ATP; phosphotransferase

Query Match 2.9%; Score 10; DB 2; Length 348;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 273 SSSSAPASSS 282
Db 46 SSSSAPASSS 55

RESULT 3
T51340
mitogen-activated protein kinase kinase (EC 2.7.1.-) 5 [imported] - Arabidopsis thaliana
N:Alternate names: MAP kinase kinase 5
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 08-Sep-2000
C:Accession: T51340
R:Ichimura, K.; Mizoguchi, T.; Hayashida, N.; Seki, M.; Shinzaki, K.
DNA Res. 5, 341-348, 1998
A:Title: Molecular cloning and characterization of three cDNAs encoding putative mitogen
A:Reference number: 225272; MUID:99156228
A:Accession: T51340
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-348 <ICH>
A:Cross-references: EMBL:AB015316; PIDN:BAA28831.1
C:Genetics:
A:Gene: ATKMK5
C:Function:
A:Description: (EC 2.7.1.-): mitogen-activated protein kinase kinase [validated; MUID:99
C:Superfamily: kinase-related transforming protein; protein kinase homology
C:Keywords: phosphotransferase; protein kinase

Query Match 2.9%; Score 10; DB 2; Length 348;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 273 SSSSAPASSS 282
Db 46 SSSSAPASSS 55

RESULT 4
S59632
endo-1,4-beta-xylanase (EC 3.2.1.8) B precursor - Cellvibrio mixtus
C:Species: Cellvibrio mixtus
C>Date: 15-Feb-1996 #sequence_revision 11-Apr-1997 #text_change 18-Jun-1999
C:Accession: S59632; S52742
R:Millward-Sadler, S.J.; Davidson, K.; Hazlewood, G.P.; Black, G.W.; Gilbert, H.J.; Cla
Biochem. J. 312, 39-48, 1995
A:Title: Novel cellulose-binding domains, Nodb homologues and conserved modular archite
A:Reference number: S59631; MUID:96077124
A:Accession: S59632
A:Molecule type: DNA

A:Residues: 1-621 <MIL>
A:Cross-references: EMBL:Z48926; NID:g757808; PIDN:CAA8762.1; PID:g757809
C:Genetics:
A:Gene: xynB
C:Function:
A:Description: catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in xylans
A:Pathway: xylan degradation
C:Superfamily: Pseudomonas endo-1,4-beta-xylanase F; Streptomyces endo-1,4-beta-xylan
F:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-621/Product: endo-1,4-beta-xylanase B #status predicted <MAT>
F:302-615/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>
F:403,516/Active site: Glu #status predicted

Query Match 2.9%; Score 10; DB 1; Length 621;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 274 SSSAPASSS 283
Db 89 SSSAPASSS 98

RESULT 5
B84259
hypothetical protein Vng1026h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84259
R:N, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leitauer, B.; Keller, K.; Cruv, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: AB4160; MUID:20504483
A:Accession: B84259
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-184 <STO>
A:Cross-references: GB:AE004437; NID:g10580580; PIDN:AAG19438.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG1026H

Query Match 2.6%; Score 9; DB 2; Length 184;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 LAVAAAAA 20
Db 137 LAVAAAAA 145

RESULT 6
S06956
segmentation protein hairy - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 21-Jul-2000
C:Accession: S06956; S06700
R:Rushlow, C.A.; Hogan, A.; Plinchin, S.M.; Howe, K.M.; Lardelli, M.; Ish-Horowitz, D.
EMBO J. 8, 3095-3103, 1989
A:Title: The Drosophila hairy protein acts in both segmentation and bristle pattern
A:Reference number: S06956; MUID:90059856
A:Accession: S06956
A:Molecule type: DNA
A:Residues: 1-337 <RUS>
A:Cross-references: GB:X15904; GB:S63792; GB:X16632; NID:g8048; PIDN:CAA34018.1; PID:
A:Note: 292-ser was also found
C:Genetics:
A:Gene: hairy
A:Cross-references: FlyBase:FBgn0001168

A:Map position: 66D
A:introns: 33/3; 65/3
C:Keywords: DNA binding; transcription regulation

Query Match 2.6%; Score 9; DB 2; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 LAVAAAAA 20
|||||
Db 238 LAVAAAAA 246

RESULT 7
A72599
hypotheetical protein APEI257 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: A72599

R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6: 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Creomarchaeon, Aeropyrum pernix
A:Reference number: A72450; MUID:99310339

A:Accession: A72599
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-439 <KAM>
A:Cross-references: DDBJ:AP000061; NID:g5104821; PIDN:BA80247.1; PID:dl044033; PID:g5104821
A:Experimental source: strain K1
C:Genetics:
A:Gene: APEI257

Query Match 2.6%; Score 9; DB 2; Length 439;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 SALAVAAAA 18
|||||
Db 398 SALAVAAAA 406

RESULT 8
A46600
methylmalonate-semialdehyde dehydrogenase (acylating) (EC 1.2.1.27) precursor - bovine

C:Species: Bos primigenius taurus (cattle)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 05-May-2000
C:Accession: A46600

R:Deichalte, I.; Berthiaume, L.; Pesceckis, S.M.; Patton, W.F.; Resh, M.D.
J. Biol. Chem. 268, 13738-13747, 1993
A:Title: Novel use of an Iodo-Myristyl-CoA analog identifies a semialdehyde dehydrogenase

A:Reference number: A46600; MUID:93293905
A:Accession: A46600
A:Status: preliminary
A:Molecule type: mRNA

A:Residues: 1-537 <DEI>
A:Cross-references: GB:L08643; NID:g289441; PIDN:AAA30650.1; PID:g289442
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: coenzyme A; mitochondrion; oxidoreductase
F:80-336/Domain: aldehyde dehydrogenase homology <ALDD>
F:319/Active site: Cys #status predicted

Query Match 2.6%; Score 9; DB 1; Length 537;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 AVAAAAAAL 21
|||||
Db 5 AVAAAAAAL 13

RESULT 9
T29064
hyaluronate lyase homolog - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: T29064
R:Redenbach, M.; Kleiser, H.M.; Denapalte, D.; Eichner, A.; Cullum, J.; Kinashi, H.; H Mol. Microbiol. 21, 77-96, 1996
A:Title: A set of ordered cosmid and a detailed genetic and physical map for the 8 M

A:Reference number: Z20556; MUID:97000351
A:Accession: T29064
A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA
A:Residues: 1-776 <RED>
A:Cross-references: EMBL:AL031124; NID:el312893; PID:el312908; PIDN:CA19982.1
C:Genetics:
A:Note: SCIC2.15

Query Match 2.6%; Score 9; DB 2; Length 776;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 AAAAAAALT 23
|||||
Db 18 AAAAAAALT 26

RESULT 10
I48719
protein kinase C (EC 2.7.1.1) mu precursor - mouse

N:Alternate names: protein kinase D
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 11-Jun-1999
C:Accession: I48719

R:Valverde, A.M.; Sinnett-Smith, J.; Van Lint, J.; Rozenfurt, E.
Proc. Natl. Acad. Sci. U.S.A. 91, 8572-8576, 1994
A:Title: Molecular cloning and characterization of protein kinase D: a target for dia

A:Reference number: I48719; MUID:94359973
A:Accession: I48719
A:Status: translated from GB/EMBL/DDBJ

A:Molecule type: mRNA
A:Residues: 1-918 <RES>
A:Cross-references: EMBL:Z34524; NID:g520877; PIDN:CA84283.1; PID:g520878

C:Function: catalyzes the formation of peptidyl-serine-phosphate or peptidyl-threo
A:Note: expressed at low levels in a variety of tissues; phosphorylates the peptide s
sely related human enzyme, this protein is reported to bind phorbol esters
C:Superfamily: protein kinase C mu; protein kinase C zeta-binding repeat homology; pr

C:Keywords: ATP; autophosphorylation; duplication; phorbol ester binding; phospholipid
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-918/Product: protein kinase mu #status predicted <MAT>

F:145-194/Domain: protein kinase C zinc-binding repeat homology <KZ1>
F:277-326/Domain: protein kinase C zinc-binding repeat homology <KZ2>
F:587-845/Domain: protein kinase C zinc-binding repeat homology <KZ3>

F:595-603/Region: protein kinase homology <KIN>
F:618,636,712,714/Active site: Lys, Glu, Asp, Lys #status predicted

Query Match 2.6%; Score 9; DB 1; Length 918;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 AVAAAAAAL 21
|||||
Db 19 AVAAAAAAL 27

RESULT 11
S42718
nuclear pore complex protein nup153 - human
C:Species: Homo sapiens (man)

C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C:Accession: S42718; S37477
R:McMorrow, I.; Bastos, R.; Horton, H.; Burke, B.
Biochim. Biophys. Acta 1217, 219-223, 1994
A:Title: Sequence analysis of a cDNA encoding a human nuclear pore complex protein, hnnpf
A:Reference number: S42718; MUID:94154002
A:Molecule type: mRNA
A:Residues: 1-1475 <MCNC>
A:Cross-references: EMBL:Z25535; NID:g406224; PIDN:CAA80982.1; PID:g406225

Query Match 2.3%; Score 8; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 282 SSVASSSS 290
|||||
DB 823 SSVASSSS 831

RESULT 12

FDPL3W
antifreeze protein 3 - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C>Date: 01-Sep-1981 #sequence_revision 01-Sep-1981 #text_change 23-Aug-1996
C:Accession: A03192
R:Devries, A.L.; Lilo, Y.
Biochim. Biophys. Acta 495, 388-392, 1977
A:Title: Structure of a peptide antifreeze and mechanism of adsorption to ice.
A:Reference number: A03192; MUID:78060969
A:Accession: A03192
A:Molecule type: protein
A:Residues: 1-37 <DEW>
C:Superfamily: antifreeze protein
C:Keywords: antifreeze

Query Match 2.3%; Score 8; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 AAAAAALT 22
|||||
DB 6 AAAAAALT 13

RESULT 13

PN0589
tyrosine 3-monooxygenase (EC 1.14.16.2) type 1 and 2 - chimpanzee (fragment)
N:Alternate names: oxygen oxidoreductase; tetrahydropteridine; tyrosine hydroxylase
C:Species: Pan troglodytes (chimpanzee)
C>Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 31-Mar-2000
C:Accession: PN0589
R:Ichinose, H.; Ohye, T.; Fujita, K.; Yoshida, M.; Ueda, S.; Nagatsu, T.
Biochem. Biophys. Res. Commun. 195, 158-165, 1993
A:Title: Increased heterogeneity of tyrosine hydroxylase in humans.
A:Reference number: PN0575; MUID:93371398
A:Accession: PN0589
A:Molecule type: genomic RNA
A:Residues: 1-45 <ICH>
A:Cross-references: GB:L14791
A:Experimental source: lymphocytes of peripheral blood
C:Comment: This enzyme catalyzes the first and rate-limiting step of catecholamine biosynthesis
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bioperin; monoxygenase; oxidoreductase

Query Match 2.3%; Score 8; DB 2; Length 45;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 AVAAAAAA 20

DB 22 AVAAAAAA 29
|||||

RESULT 14

PN0590
tyrosine 3-monooxygenase (EC 1.14.16.2) type 1 and 2 - gorilla (fragment)
N:Alternate names: oxygen oxidoreductase; tetrahydropteridine; tyrosine hydroxylase
C:Species: Gorilla gorilla (gorilla)
C>Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 31-Mar-2000
C:Accession: PN0590
R:Ichinose, H.; Ohye, T.; Fujita, K.; Yoshida, M.; Ueda, S.; Nagatsu, T.
Biochem. Biophys. Res. Commun. 195, 158-165, 1993
A:Title: Increased heterogeneity of tyrosine hydroxylase in humans.
A:Reference number: PN0575; MUID:93371398
A:Accession: PN0590
A:Molecule type: genomic RNA
A:Residues: 1-45 <ICH>
A:Cross-references: GB:L14797
A:Experimental source: lymphocytes of peripheral blood
C:Comment: This enzyme catalyzes the first and rate-limiting step of catecholamine biosynthesis
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bioperin; monoxygenase; oxidoreductase

Query Match 2.3%; Score 8; DB 2; Length 45;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 AVAAAAAA 20
|||||
DB 22 AVAAAAAA 29

RESULT 15

PN0591
tyrosine 3-monooxygenase (EC 1.14.16.2) type 1 and 2 - orangutan (fragment)
N:Alternate names: oxygen oxidoreductase; tetrahydropteridine; tyrosine hydroxylase
C:Species: Pongo pygmaeus (orangutan)
C>Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 31-Mar-2000
C:Accession: PN0591
R:Ichinose, H.; Ohye, T.; Fujita, K.; Yoshida, M.; Ueda, S.; Nagatsu, T.
Biochem. Biophys. Res. Commun. 195, 158-165, 1993
A:Title: Increased heterogeneity of tyrosine hydroxylase in humans.
A:Reference number: PN0575; MUID:93371398
A:Accession: PN0591
A:Molecule type: genomic RNA
A:Residues: 1-45 <ICH>
A:Cross-references: GB:L14800
A:Experimental source: lymphocytes of peripheral blood
C:Comment: This enzyme catalyzes the first and rate-limiting step of catecholamine biosynthesis
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bioperin; monoxygenase; oxidoreductase

Query Match 2.3%; Score 8; DB 2; Length 45;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 AVAAAAAA 20
|||||
DB 22 AVAAAAAA 29

Search completed: August 23, 2001, 11:18:05
Job time: 212 sec

Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VSKADSGALTYLLEEVYKFEARMKMAASGVSMFLYONCSFLADSPVENVIEV 61
 DB 25 VSKADSGALTYLLEEVYKFEARMKMAASGVSMFLYONCSFLADSPVENVIEV 84

QY 62 LGNPGSFQSNITTKGAKQKTEKHHAVSPADQAFHTGLTEPNVYVMTVDGQEVK 121
 DB 85 LGNPGSFQSNITTKGAKQKTEKHHAVSPADQAFHTGLTEPNVYVMTVDGQEVK 144

QY 122 TEGGVSNLTGTGGLRNLMSSSAAMVGOFPDSKLPFQFINWVKYKTKPGGEGGSD 181
 DB 145 TEGGVSNLTGTGGLRNLMSSSAAMVGOFPDSKLPFQFINWVKYKTKPGGEGGSD 204

QY 182 FTLDWTFNFTFGSGRKGDMFTDGNRVDLTJKNITSRDGMILLATRKGGSPNGQVP 241
 DB 205 FTLDWTFNFTFGSGRKGDMFTDGNRVDLTJKNITSRDGMILLATRKGGSPNGQVP 264

QY 242 RDDEPAP 248
 DB 265 RDDEPAP 271

RESULT 2
 CYSO_RHIME STANDARD; PRT; 317 AA.
 ID CYSO_RHIME STANDARD; PRT; 317 AA.
 AC P56892;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SULFATE ADENYLTRANSFERASE SUBUNIT 2 (EC 2.7.7.4) (SULFATE ADENYLATE TRANSFERASE) (SAT) (ATP-SULFURYLASE SMALL SUBUNIT).
 GN CYSO.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=99395034; PubMed=10464198;
 RA Abola A.P., Willits M.G., Wang R.C., Long S.R.;
 RT "Reduction of adenosine-5'-phosphosulfate instead of 3'-phosphodenosine-5'-phosphosulfate in cysteine biosynthesis by Rhizobium meliloti and other members of the family Rhizobiaceae.";
 RL J. Bacteriol. 181:5280-5287(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA Gallbert F., Capela D., Hubler-Barloy F., Galius M., Batut J.,
 RA Boistard P., Gouzy J., Kahn D., Thebaud P., Goffeau A.,
 RA Purrelle B., Pohl T., Bothe G., Schneider S., Portetelle D.,
 RA Vandenbol M., Puehler A., Becker A., Weidner S.;
 RL Submitted (May-2000) to the SWISS-PROT data bank.
 CC -1- CATALYTIC ACTIVITY: ATP + SULFATE -> PYROPHOSPHATE + ADENYLSULFATE.
 CC -1- PATHWAY: FIRST STEP IN THE SULFATE ACTIVATION PATHWAY. THIS REACTION OCCURS EARLY IN THE REDUCTIVE BRANCH OF THE CYSTEINE BIOSYNTHETIC PATHWAY.
 CC -1- SUBUNIT: HETERODIMER COMPOSED OF CYSO, THE SMALLER SUBUNIT, AND CYSN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY. CYSO SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: AF158023; AAD55760.1; -
 DR InterPro: IPR002500; -

DR Pfam: PF01507; PAPS_reduct; 1.
 KM Cysteine biosynthesis; Transferase; Nucleotidyltransferase.
 FT CONFLICT 260
 SQ SEQUENCE 317 AA; 36429 MM; 093F2724D22841CB CRC64;

Query Match 3.0%; Score 8; DB 1; Length 317;
 Best local similarity 100.0%; Pred. No. 1.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 RDGMILLA 227
 DB 234 RDGMILLA 241

RESULT 3
 F16P_PEA STANDARD; PRT; 407 AA.
 ID F16P_PEA STANDARD; PRT; 407 AA.
 AC P46275; Q37263;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE FRUCTOSE-1,6-BISPHOSPHATASE, CHLOROPLAST PRECURSOR (EC 3.1.3.11)
 DE (D-FRUCTOSE-1,6-BISPHOSPHATE 1-PHOSPHOHYDROLASE) (FBPase).
 GN FBP.
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 OC Fabales; Fabaceae; Papilionoideae; Pisum.
 OX NCBI_TaxID=3888;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. GIANT; TISSUE=leaf;
 RX MEDLINE=95175626; PubMed=7870839;
 RA Dong S.M., Rhim J.H., Hahn T.R.;
 RT "Nucleotide sequence analysis of a cDNA encoding chloroplastic fructose-1,6-bisphosphatase from pea (Pisum sativum L.).";
 RL Plant Physiol. 107:313-314(1995).
 RN [2]
 RP SEQUENCE OF 27-407 FROM N.A.
 RC STRAIN=CV. LINCOLN; TISSUE=leaf;
 RX MEDLINE=94297517; PubMed=7764999;
 RA Carrasco J.L., Chuca A., Prado F.E., Hermoso R., Lazaro J.J.,
 RA Ramos J.L., Sahrawy M., Lopez Gorge J.;
 RT "Cloning, structure and expression of a pea cDNA clone coding for a photosynthetic fructose-1,6-bisphosphatase with some features different from those of the leaf chloroplast enzyme.";
 RL Planta 193:494-501(1994).
 CC -1- CATALYTIC ACTIVITY: D-FRUCTOSE 1,6-BISPHOSPHATE + H(2O) -> D-FRUCTOSE 6-PHOSPHATE + ORTHOPHOSPHATE.
 CC -1- PATHWAY: THE CHLOROPLAST ISOZYME TAKES PART IN THE REGENERATION OF RIBULOSE BISPHOSPHATE IN THE PHOTOSYNTHETIC CARBON REDUCTION CYCLE (CALVIN CYCLE).
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST STROMA.
 CC -1- INDUCTION: LIGHT ACTIVATION THROUGH PH CHANGES, MG(2+) LEVELS AND ALSO BY LIGHT-MODULATED REDUCTION OF ESSENTIAL DISULFIDE GROUPS VIA THE FERREDOXIN-THIOREDOXIN F SYSTEM (BY SIMILARITY).
 CC -1- MISCELLANEOUS: IN PLANTS THERE ARE TWO FBPASE ISOZYMES: ONE IN THE CYTOSOL AND THE OTHER IN THE CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE FBPASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: U34806; AAD10213.1; -
 DR EMBL: X68826; CAA48719.1; -
 DR HSP; P22418; 1SP1.

```

DR InterPro: IPR000146; -
DR Pfam; PF00316; FBpase; 1.
DR PRINTS; PR00115; FBpPHTASE.
DR PRINTS; PR00377; INBPpHTASE.
DR PROSITE; PS00124; FBpase; 1.
KM Hydrolyase; Carbohydrate metabolism; Multigene family; Chloroplast;
KW Transit peptide; Calvin cycle.
FT TRANSIT 1 50
FT CHAIN 51 407
FT ACT_SITE 349 349
FT DISULFID 223 228
FT FT
FT FT
FT CONFLICT 82 82
FT CONFLICT 160 160
FT CONFLICT 247 247
FT CONFLICT 282 282
SQ SEQUENCE 407 AA; 44511 MW; B60E9164F1FEFEDD CRC64;

Query Match
Best Local Similarity 3.0%; Score 8; DB 1; Length 407;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 128 SMLGTG 135
DB 112 SMLGTG 119

RESULT 4
OCRA_MYCTU STANDARD; PRT; 429 AA.
ID OCRA_MYCTU
AC Q10387;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE UBIQUINOL-CYTOCHROME C REDUCTASE IRON-SULFUR SUBUNIT (RIESKE IRON-
DE SULFUR PROTEIN).
GN OCRA OR RV2195 OR MTCY190.06.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; Pubmed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX. THE RIESKE PROTEIN IS A HIGH POTENTIAL 2FE-2S PROTEIN.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF THE UBIQUINOL-CYTOCHROME C COMPLEX
CC ARE: CYTOCHROME B, THE RIESKE PROTEIN AND A CYTOCHROME C SUBUNIT.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO RIESKE PROTEINS FROM OTHER SOURCES (MITOCHONDRIA,
CC BACTERIAL, CHLOROPLAST).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL; Z70283; CAA94264.1; -
DR TubercList; RV2195; -
DR InterPro; IPR001281; -
DR PROSITE; PS00199; RIESKE_1; FALSE_NEG.
DR PROSITE; PS00200; RIESKE_2; 1.
KW Electron transport; Iron-sulfur; Oxidoreductase; Transmembrane.
FT TRANSMEM 96 116
FT TRANSMEM 137 157
FT TRANSMEM 207 227
FT METAL 353 353
FT METAL 355 355
FT METAL 372 372
FT METAL 375 375
FT DISULFID 358 374
SQ SEQUENCE 429 AA; 46922 MW; C22C87EC283FE0EE CRC64;

Query Match
Best Local Similarity 3.0%; Score 8; DB 1; Length 429;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 230 RKGOESFN 237
DB 335 RKGOESFN 342

RESULT 5
HPN_HELPY STANDARD; PRT; 59 AA.
ID HPN_HELPY
AC Q48251;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HISTIDINE-RICH, METAL BINDING POLYPEPTIDE.
GN HPN OR HP1427 OR JHP1320.
OS Helicobacter pylori (Campylobacter pylori), and
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210, 85963;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-18 AND 47-59.
RC STRAIN=LEU;
RX MEDLINE=95310028; Pubmed=7790085;
RA Gilbert J.V., Ramakrishna J., Sunderman F.W. Jr., Wright A.,
RA Plant A.G.;
RT "Protein Hpn: cloning and characterization of a histidine-rich metal-
RT binding polypeptide in Helicobacter pylori and Helicobacter
RT mustelae."
RL Infect. Immun. 63:2682-2688(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; Pubmed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Bordovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 386:539-547(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J99;
RX MEDLINE=99120557; Pubmed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deGange B.L., Carmel G.,
RA Tummlino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,

```

RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen *Helicobacter pylori*.";
 RL Nature 397:176-180(1999).
 CC -1- FUNCTION: STRONGLY BINDS NICKEL AND ZINC. BINDS OTHER METALS LESS
 CC STRONGLY: CO2+ > CU2+ > CD2+ > MN2+. MAY ACT TO INCREASE, OR AT
 CC LEAST TO PRESERVE, UREASE ACTIVITY. EXACT FUNCTION IS STILL
 CC UNKNOWN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U26361; AAA5859.1; -
 CC EMBL: AE000643; AAD08471.1; -
 CC EMBL: AE001555; AAD06898.1; -
 CC TIGR: HP1427; -
 CC KM Metal-binding: Zinc; Nickel; Repeat.
 CC INIT_MER 0 0
 CC FT DOMAIN 10 23 POLY-HIS.
 CC FT DOMAIN 27 32 POLY-HIS.
 CC FT DOMAIN 37 54 2 X 5 AA REPEATS OF E-E-G-C-C.
 CC FT REPEAT 37 41 1.
 CC FT REPEAT 50 54 2.
 CC FT REPEAT 59 AA; 6946 MW; C3AE3F602EC973C CRC64;
 CC SEQUENCE

Query Match 2.6%; Score 7; DB 1; Length 59;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 EHHNNH 267
 DB 26 EHHNNH 32

RESULT 6
 ID YWKF_BACSU STANDARD; PRT; 95 AA.
 AC P45874;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE HYPOTHETICAL 10.1 KDA PROTEIN IN PFAA-SPOTIR INTERGENIC REGION.
 GN YWKF.
 OS *Bacillus subtilis*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC *Bacillus/Staphylococcus* group; *Bacillus*.
 CX NCBI_Taxid-1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-168;
 RC STRAIN-168;
 RL Glaser P., Danchin A.;
 RL Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 75-95 FROM N.A.
 CC -----
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC -----
 CC EMBL: Z49782; CAAB9886.1; -
 CC EMBL: Z38002; -; NOT ANNOTATED_CDS.
 CC EMBL: Z99122; CAB15716.1; -
 CC DR Subtilisin; BG11317; ywKF.
 CC KW Hypothetical protein; Transmembrane.
 CC FT TRANSMEM 5 25 POTENTIAL.
 CC FT TRANSMEM 31 51 POTENTIAL.
 CC FT TRANSMEM 75 95 POTENTIAL.
 CC SEQUENCE 95 AA; 10069 MW; 2A2F80BDEC705BC4 CRC64;
 CC -----

Query Match 2.6%; Score 7; DB 1; Length 95;
 Best Local Similarity 100.0%; Pred. No. 6.8;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RMKAAA 32
 DB 71 RMKAAA 77

RESULT 7
 ID YGH2_YEAST STANDARD; PRT; 119 AA.
 AC P53161;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHETICAL 13.4 KDA PROTEIN IN HSF1-AFT1 INTERGENIC REGION.
 GN YGH2.
 OS *Saccharomyces cerevisiae* (Baker's yeast).
 CC *Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;*
 CC *Saccharomycetales; Saccharomycetaceae; Saccharomyces*.
 CX NCBI_Taxid-4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-5288C;
 RX MEDLINE-97435481; PubMed-9290212;
 RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
 RT "Sequence analysis of 203 kilobases from *Saccharomyces cerevisiae*
 RT chromosome VII.";
 RL Yeast 13:1077-1090(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z72596; CAA96778.1; -
 CC DR SGD: S0003040; YGL072C.
 CC KW Hypothetical protein; Transmembrane.
 CC FT TRANSMEM 30 50 POTENTIAL.
 CC SEQUENCE 119 AA; 13388 MW; F1D2F040C1C83A6 CRC64;
 CC -----

Query Match 2.6%; Score 7; DB 1; Length 119;
 Best Local Similarity 100.0%; Pred. No. 8.3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 FSGAELY 13
 DB 98 FSGAELY 104

RESULT 8
 ID NODP_RHS3 STANDARD; PRT; 301 AA.
 AC P72338;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SULFATE ADENYLYLTRANSFERASE SUBUNIT 2 (EC 2.7.7.4) (SULFATE ADENYLYLTRANSFERASE) (SAT) (ATP-SULFURYLASE SMALL SUBUNIT) (MODULATION PROTEIN P.).
DE P.).
DE GN
OS Rhizobium sp. (strain N33).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=103798;
RN [1]
RP SEQUENCE FROM N.A.
RA Cloutier J.;
RL Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PROPOSED TO PROVIDE ACTIVATED SULFATE FOR TRANSFER TO NOD FACTOR.
CC -1- CATALYTIC ACTIVITY: ATP + SULFATE = PYROPHOSPHATE + ADENYLYLSULFATE.
CC -1- SUBUNIT: SULFATE-ACTIVATING ENZYMES, NODP AND NODO, MAY BE PHYSICALLY ASSOCIATED (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY, CYSD SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U53327; AAB16901.1; -.
DR InterPro; IPR002500; -.
DR Pfam; PF01507; PAPS_reduct; 1.
KW Modulation; Transferase; Nucleotidyltransferase.
SQ SEQUENCE 301 AA; 35008 MW; 3582F7EAB0861431 CMC64;

Query Match 2.6%; Score 7; DB 1; Length 301;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 RDGMIL 226
|111111|
DB 218 RDGMIL 224

RESULT 9
HEMZ_BRAJA STANDARD; PRT; 345 AA.
AC P28602;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE FERROCHELATASE (EC 4.99.1.1) (PROTOHEME FERRO-LYASE) (HEME SYNTHETASE).
DE HEMH.
GN Bradyrhizobium japonicum.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LO;
RX MEDLINE=92325004; Pubmed=1624416;
RA Frustaci J.M., O'Brian M.R.;
RT "Characterization of a Bradyrhizobium japonicum ferrochelatase mutant and isolation of the hemh gene.";
RL J. Bacteriol. 174:4223-4229(1992).
RN [2]
RP REVISIONS.
RA O'Brian M.;
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE FERROUS INSERTION INTO PROTOPORPHYRIN IX. ESSENTIAL FOR NORMAL NODULE DEVELOPMENT.

CC -1- CATALYTIC ACTIVITY: PROTOPORPHYRIN + FE(2+) = PROTOHEME + 2 H(+).
CC -1- PATHWAY: LAST STEP IN PROTOHEME BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FERROCHELATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M92427; AAA6217.2; -.
DR PIR; A42883; A42883.
DR InterPro; IPR001015; -.
DR Pfam; PF00762; Ferrochelatase; 1.
DR PROSITE; PS00534; FERROCHELATASE; 1.
KW Porphyrin biosynthesis; Heme biosynthesis; Lyase; Iron.
FT METAL 215 215 IRON (BY SIMILARITY).
FT METAL 296 296 IRON (BY SIMILARITY).
SQ SEQUENCE 345 AA; 38431 MW; 23E1658F40D38788 CMC64;

Query Match 2.6%; Score 7; DB 1; Length 345;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 DKLAAL 260
|111111|
DB 102 DKLAAL 108

RESULT 10
HREX_PLALO STANDARD; PRT; 351 AA.
ID HREX_PLALO
AC P04929;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1989 (Rel. 38, Last annotation update)
DE HISTIDINE-RICH GLYCOPROTEIN PRECURSOR.
OS Plasmodium lophurae.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5853;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85061618; Pubmed=6095114;
RA Ravetch J.V., Feder R., Pavlovic A., Blobel G.;
RT "Primary structure and genomic organization of the histidine-rich protein of the malaria parasite Plasmodium lophurae.";
RL Nature 312:616-620(1984).
CC -1- MISCELLANEOUS: IN THE INTRAERYTHROCYTIC STAGES OF DEVELOPMENT OF P. LOPHURAE IN DUCKS, THERE IS A SYNTHESIS OF A MAJOR PROTEIN THAT ACCUMULATES TO COMPRISE AT LEAST 50% OF THE CELLULAR MASS: THE HISTIDINE RICH PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X01469; CAA25698.1; -.
DR PIR; A22692; KGZOH.
KW Malaria; Repeat; Glycoprotein; Signal.
FT SIGNAL 1 23
FT PROPEP 24 47
FT CHAIN 48 351
FT CARBOHD 40 40
FT DOMAIN 59 90
FT REPEAT 59 74
HISTIDINE-RICH GLYCOPROTEIN
N-LINKED (GLCNAC...) (PROBABLE).
2 X 16 AA TANDEM REPEATS.
16-1.

FT REPEAT 75 90 16-2.
 FT DOMAIN 91 123 2 x 17 AA TANDEM REPEATS.
 FT REPEAT 91 107 17-1.
 FT DOMAIN 108 123 17-2.
 FT REPEAT 124 153 2 x 15 AA TANDEM REPEATS.
 FT REPEAT 124 138 15-1.
 FT REPEAT 139 153 15-2.
 FT DOMAIN 173 351 18 x 10 AA TANDEM REPEATS.
 SQ SEQUENCE 351 AA; 44032 MW; D19A48D47D890453 CRC64;

Query Match 2.6%; Score 7; DB 1; Length 351;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 261 EHHHHH 267
 DB 132 EHHHHH 138

RESULT 11
 MIX2_XENLA STANDARD; PRT; 369 AA.
 ID MIX2_XENLA
 AC 091685;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HOMEOBOX PROTEIN MIX.2.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae;
 OC Xenopodidae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9629435; Pubmed=8660890;
 RA Vize P.D.;
 RT "DNA sequences mediating the transcriptional response of the Mix.2 homeobox gene to mesoderm induction."
 RL Dev. Biol. 177:226-231(1996).
 CC -1- FUNCTION: MAY BE A TRANSCRIPTION FACTOR WHICH PLAY A REGULATORY ROLE IN THE DEVELOPMENT OF THE EMBRYO (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- INDUCTION: BY ACTIVIN AND TGF-BETA1 (IMMEDIATE EARLY RESPONSE GENE).
 CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEOBOX PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: 050745; AAC60020.1; -
 CC DR HSSP: P06601; 1FJL.
 CC DR InterPro: IPR001356; -
 CC DR Pfam: PF00046; homeobox.1;
 CC DR PROSITE: PS00027; HOMEOBOX_1;
 CC DR PROSITE: PS0071; HOMEOBOX_2; 1.
 CC KW Homeobox: DNA-binding; Developmental protein; Nuclear protein.
 CC FT DNA_BIND 94 153 HOMEOBOX.
 CC SQ SEQUENCE 369 AA; 41537 MW; FC02182C85D6B9F5 CRC64;

Query Match 2.6%; Score 7; DB 1; Length 369;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 246 PAPNSSS 252
 DB 178 PAPNSSS 184

RESULT 12
 MIX1_XENLA STANDARD; PRT; 377 AA.
 ID MIX1_XENLA
 AC P21711;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HOMEOBOX PROTEIN MIX.1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae;
 OC Xenopodidae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89288302; Pubmed=2567635;
 RA Rosa F.M.;
 RT "Mix.1, a homeobox mRNA inducible by mesoderm inducers, is expressed mostly in the presumptive endodermal cells of Xenopus embryos."
 RL Cell 57:965-974(1989).
 CC -1- FUNCTION: MAY BE A TRANSCRIPTION FACTOR WHICH PLAY A REGULATORY ROLE IN THE DEVELOPMENT OF THE EMBRYO. INVOLVED IN THE ESTABLISHMENT OF DORSAL/VENTRAL PATTERN IN THE EARLY MESODERM.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- TISSUE SPECIFICITY: EXPRESSED MOSTLY IN THE PRESUMPTIVE ENDODERMAL CELLS OF XENOPUS EMBRYOS.
 CC -1- INDUCTION: BY ACTIVIN AND TGF-BETA1 (IMMEDIATE EARLY RESPONSE GENE).
 CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEOBOX PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: M21063; AAA49903.1; -
 CC DR PIR: A32548; A32548.
 CC DR HSSP: P06601; 1FJL.
 CC DR InterPro: IPR001356; -
 CC DR Pfam: PF00046; homeobox.1;
 CC DR PROSITE: PS00027; HOMEOBOX_1;
 CC DR PROSITE: PS0071; HOMEOBOX_2; 1.
 CC KW Homeobox: DNA-binding; Developmental protein; Nuclear protein.
 CC FT DNA_BIND 96 155 HOMEOBOX.
 CC FT DOMAIN 352 368 ASP/GLU-RICH (ACIDIC).
 CC SQ SEQUENCE 377 AA; 42250 MW; 6BA404359B57D221 CRC64;

Query Match 2.6%; Score 7; DB 1; Length 377;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 246 PAPNSSS 252
 DB 183 PAPNSSS 189

RESULT 13
 HMPA_ECOLI STANDARD; PRT; 396 AA.
 ID HMPA_ECOLI
 AC P24232;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE FLAVOHEMOPROTEIN (HEMOGLOBIN-LIKE PROTEIN) (FLAVOHEMOGLOBIN) (DIHYDROPTERIDINE REDUCTASE (EC 1.6.99.7)) (FERRISIDEROPHORE REDUCTASE B) (NITRIC OXIDE DIOXYGENASE) (NOD).
 DE HMP OR HMPA OR FSRB.

OS Escherichia coli.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
OC Escherichia
CX NCBI_Taxid=562;
RN (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.
RC STRAIN-K12;
RC MEDLINE=91238719; PubMed=2034230;
RA Vasudevan S.G., Armarego W.L.F., Shaw D.C., Lilley P.E., Dixon N.E.,
RA Poole R.K.;
RT "Isolation and nucleotide sequence of the hmp gene that encodes a
RT haemoglobin-like protein in Escherichia coli K-12";
RL Mol. Gen. Genet. 226:49-58(1991).
RN (12)
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RC MEDLINE=97456617; PubMed=9278503;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horikuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RN DNA Res. 4:91-113(1997).
RN (14)
RP SEQUENCE OF 1-10 FROM N.A.
RC MEDLINE=83235562; PubMed=6190704;
RA Pliamam M.D., Stauffer G.V.;
RT "Characterization of the Escherichia coli gene for serine
RT hydroxymethyltransferase.";
RL Gene 22:9-18(1983).
RN (15)
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC STRAIN-K12;
RC MEDLINE=92290008; PubMed=1601132;
RA Andrews S.C., Shipley D., Keen J.N., Findlay J.B.C., Harrison P.M.,
RA Guest J.R.;
RT "The haemoglobin-like protein (HMP) of Escherichia coli has
RT ferrisulphophore reductase activity and its C-terminal domain shares
RT homology with ferredoxin NADP+ reductases.";
RL FEBS Lett. 302:247-252(1992).
RN (16)
RP SEQUENCE OF 1-12.
RC STRAIN-K12 / EWG2;
RC MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12";
RL Electrophoresis 18:11259-1311(1997).
RN (17)
RP SEQUENCE OF 1-20, AND CHARACTERIZATION.
RC MEDLINE=98393652; PubMed=9724711;
RA Gardner P.R., Gardner A.M., Martin L.A., Salzman A.L.;
RT "Nitric oxide dioxygenase: an enzymic function for flavohaemoglobin.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:10378-10383(1998).
CC -1- FUNCTION: POSSSSESSES DIHYDROPTERIDINE REDUCTASE ACTIVITY. ALSO HAS
CC NITRIC OXIDE DIOXYGENASE ACTIVITY.
CC -1- FUNCTION: SEEMS TO REDUCES THE COMPLEXED FE+3 IRON OF SIDEROPHORES
CC TO FE+2. THUS RELEASING IT FROM THE CHELATOR.
CC -1- CATALYTIC ACTIVITY: NAD(P)H + 6,7-DIHYDROPTERIDINE = NAD(P)(+) +

```
CC      5, 6,7,8-TETRAHYDROPTERIDINE.
CC      -1- SUBUNIT: MONOMER..
CC      -1- DOMAIN: CONSISTS OF TWO DISTINCT DOMAINS; ONE IS A HEME-CONTAINING
CC      OXYGEN BINDING DOMAIN IN THE N-TERMINAL REGION AND THE OTHER IS AN
CC      FAD-CONTAINING REDUCTASE DOMAIN FOUND IN THE C-TERMINAL REGION..
CC      -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY. TWO DOMAINS
CC      FLAVOHEMOPROTEINS SUBFAMILY.
CC      -1- SIMILARITY: TO A NUMBER OF FAD/NAD(P) FLAVOPROTEIN
CC      OXIDOREDUCTASES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X58872; CAA41682.1; -.
DR      EMBL; AE000341; AAC75605.1; -.
DR      EMBL; D90885; BA016460.1; -.
DR      EMBL; J01620; AAA23911.1; -.
DR      PIR; S15992; S15992.
DR      PIR; S21161; S21161.
DR      HSSP; P04252; 1VHB.
DR      EcoGene; EG10456; hmp.
DR      InterPro; IPR000971; -.
DR      InterPro; IPR001433; -.
DR      Pfam; PF00042; globin; 1.
DR      Pfam; PF00175; oxidored.fad; 1.
DR      PROSITE; PS01033; GLOBIN; 1.
KW      Oxidoreductase; NADP; Heme; Flavoprotein; FAD; Iron transport.
FT      METAL
FT      FT   METAL          1       134           GLOBIN.
FT      FT   METAL          53        53           IRON (HEME DISTAL LIGAND)
FT      FT   METAL          85        85           (BY SIMILARITY).
FT      FT   METAL          85        85           IRON (HEME PROXIMAL LIGAND)
FT      FT   METAL          85        85           (BY SIMILARITY)..
FT      NP_BIND      268      273      NADP (RIBOSE P4RY) (BY SIMILARITY).
SQ      SEQUENCE      396 AA; 43867 MW; 49961BDE1444BD68 CRC64;

Query Match                               2.6%; Score 7; DB 1; Length 396;
Best Local Similarity 100.0%; Pred. No. 24;
Matches    7; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

QY      123 EGGGVSN 129
       |||||
Db      225 EGGGVSN 231

RESULT 14
YDJN_HAEIN
AC      PA5079;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      HYPOTHEITICAL SYMPORFER H1154.
GN      H1154.
OS      Haemophilus influenzae.
OC      Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC      Haemophilus.
OX      NCBI_TaxID=727;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-RD / KW20 / ATCC 51907.
RX      MEDLINE=95350630; PubMed=7542800;
RA      Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA      Kelleys A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA      McEneney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA      Scott J.D., Shiley R., Liu L.-I., Glock A., Kelley J.M.,
RA      Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA      Uetlerack T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
```

```

RA Fine L.D., Fritschman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd.";
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
CC (SDF). STRONG, TO E.COLI YDJN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32795; AAC22809.1; -.
CC TIGR: H1154; -.
CC InterPro: IPR001991; -.
CC Pfam: PF00375; SDF: 1.
CC DR PROSITE: PS00713; NA_DICARBOXYL_SYM_1; FALSE_NEG.
CC DR PROSITE: PS00714; NA_DICARBOXYL_SYM_2; FALSE_NEG.
CC KW Hypothetical protein; Transmembrane; Transport.
CC FT TRANSMEM 1 21 POTENTIAL.
CC FT TRANSMEM 29 49 POTENTIAL.
CC FT TRANSMEM 70 90 POTENTIAL.
CC FT TRANSMEM 101 121 POTENTIAL.
CC FT TRANSMEM 179 199 POTENTIAL.
CC FT TRANSMEM 226 246 POTENTIAL.
CC FT TRANSMEM 258 278 POTENTIAL.
CC FT TRANSMEM 343 363 POTENTIAL.
CC FT TRANSMEM 366 386 POTENTIAL.
CC FT TRANSMEM 389 409 POTENTIAL.
CC SO SEQUENCE 440 AA; 46491 MW; F929839559PD06A CRC64;

Query Match 2.6%; Score 7; DB 1; Length 440;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 GOVSNLT 131
Db 150 GOVSNLT 156

RESULT 15
ETS2_MOUSE
ID ETS2_MOUSE STANDARD; PRT; 468 AA.
AC P15037;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE C-ETS-2 PROTEIN.
GN ETS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=89042086; PubMed=2847145;
RA Watson D.K., McWilliams M.J., Lapie P., Lautenberger J.A.,
RA Schweinfest C.W., Papas T.S.;
RT "Mammalian ets-1 and ets-2 genes encode highly conserved proteins.";
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J04103; AAA37581.1; -.
CC PIR: C32066; TVMSE2.
CC DR HSSP: P14921; 2STW.
CC DR TRANSFAC: T01397; -.
CC DR MGD: MGI:95456; Ets2.
CC DR InterPro: IPR000418; -.
CC DR Pfam: PF00178; Ets; 1.
CC DR PRINTS: PR00454; ETSDOMAIN.
CC DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
CC DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
CC DR PROSITE: PS0061; ETS_DOMAIN_3; 1.
CC KW Proto-oncogene; DNA-binding; Nuclear protein.
CC FT DOMAIN 87 170 POINTED.
CC FT DNA_BIND 362 442 ETS_DOMAIN.
CC SO SEQUENCE 468 AA; 52827 MW; 5260F3085B7EB831 CRC64;

Query Match 2.6%; Score 7; DB 1; Length 468;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 FDFPDGS 196
Db 29 FDFPDGS 35

```

Search completed: August 23, 2001, 11:18:28
Job time: 195 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:14:28 ; Search time 45.86 Seconds

(without alignments)
461.356 Million cell updates/sec

Title: US-09-654-652a-3
Perfect score: 1824
Sequence: 1 MNIKRTAVKSAALVAAAAA.....AKGAKVNPNGKRYRVEFH 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	257.5	14.1	308	AAW93001	B. alkalophilus be
2	250.5	13.7	242	AAW95000	Bacillus subtilis
3	248	13.6	239	AAW06621	Hybrid (1,3-1,4)-p
4	228	12.5	234	AAW03775	Thermostable beta-
5	228	12.5	237	AAW05803	Heat-stable endo-b
6	222.5	12.2	237	AAW06622	Hybrid (1,3-1,4)-p
7	190.5	10.4	245	AAW37884	Lichenase protein.
8	185	10.1	282	AAW48467	Arabidopsis thalia
9	173	9.5	282	AAW32464	Arabidopsis thalia
10	172.5	9.5	277	AAW52114	Arabidopsis thalia
11	170.5	9.3	269	AAW18644	Arabidopsis thalia

12	170.5	9.3	282	21	AAW18643	Arabidopsis thalia
13	170.5	9.3	298	21	AAW18642	Arabidopsis thalia
14	157.5	8.6	269	21	AAW20347	Arabidopsis thalia
15	157.5	8.6	269	21	AAW20350	Arabidopsis thalia
16	157.5	8.6	269	21	AAW52653	Arabidopsis thalia
17	157.5	8.6	290	21	AAW20346	Arabidopsis thalia
18	157.5	8.6	297	21	AAW52652	Arabidopsis thalia
19	157.5	8.6	300	21	AAW20349	Arabidopsis thalia
20	156.5	8.6	269	21	AAW20329	Arabidopsis thalia
21	156.5	8.6	299	21	AAW20328	Arabidopsis thalia
22	154.5	8.5	306	17	AAW97362	Oerskovia beta-1,3
23	154	8.4	247	20	AAW11593	Protein encoded by
24	154	8.4	247	20	AAW08308	A. thaliana merist
25	145.5	8.0	307	21	AAW40361	Arabidopsis thalia
26	145.5	8.0	310	21	AAW40360	Arabidopsis thalia
27	145.5	8.0	312	21	AAW40359	Arabidopsis thalia
28	145	7.9	263	18	AAW29455	Oerskovia xanthine
29	144.5	7.9	307	21	AAW16385	Arabidopsis thalia
30	144.5	7.9	310	21	AAW16384	Arabidopsis thalia
31	144.5	7.9	313	21	AAW16383	Arabidopsis thalia
32	144	7.9	435	18	AAW29456	Oerskovia xanthine
33	143	7.8	204	21	AAW48468	Arabidopsis thalia
34	141.5	7.8	303	18	AAW29457	Oerskovia xanthine
35	141	7.7	261	12	AAW11599	Beta-1,3-glucanase
36	140	7.7	204	21	AAW32465	Arabidopsis thalia
37	139.5	7.6	204	21	AAW52115	Arabidopsis thalia
38	133.5	7.3	291	21	AAW11993	Arabidopsis thalia
39	133.5	7.3	291	21	AAW52470	Arabidopsis thalia
40	130.5	7.2	199	21	AAW20330	Arabidopsis thalia
41	130.5	7.2	199	21	AAW20348	Arabidopsis thalia
42	130.5	7.2	199	21	AAW20351	Arabidopsis thalia
43	130.5	7.2	199	21	AAW52654	Arabidopsis thalia
44	129	7.1	158	21	AAW48033	Arabidopsis thalia
45	129	7.1	171	21	AAW48032	Arabidopsis thalia

ALIGNMENTS

RESULT 1	AAW93001	standard; Protein; 308 AA.
ID	AAW93001	standard; Protein; 308 AA.
XX	AAW93001;	
AC	19-MAY-1999	(first entry)
XX		
DT		
XX		
DE		B. alkalophilus beta-glucanase protein.
XX		
KW		Beta-glucanase; enzyme; glucan; removal; lichenan; membrane;
KW		equipment; food industry; brewing.
XX		
OS		Bacillus alkalophilus.
XX		
PN	DE19732751-A1.	
XX		
PD	04-FEB-1999.	
XX		
PF	30-JUL-1997;	97DE-1032751.
XX		
PR	30-JUL-1997;	97DE-1032751.
XX		
PA	(HENK) HENKEL KGAA.	
PI		
XX	Hillen W, Maurer K;	
XX		
DR	WPI; 1999-122161/11.	
XX	N-PSDB; AAX02912.	
PT		New Bacillus alkalophilus beta-glucanase protein and gene - useful
PT		for removing glucan and/or lichenan from membranes in the brewing
XX	industry	


```

Db      77  tspsynkfdcgensvqlygylyevrmkpkaknlgivssftlycpte---gtpwdeidi 133
QY      83  EVLGKNGSPGFSNIIITGKAGAKTSEKHHAASPAADQAFHTYGLEMTNRYRWYVDDGGEV 142
      134  eflgkdtlcvfnytlngvgqh---ekvisigfaskgfhyaftwqpyikwyvdg-vl 189
QY      143  RKTGGGVSNLTGTQG-LRFNLMSEESA-AWVGOFDESKLPFQFINWVKYKXT 195
      190  kht---atanlpsrpgkimmlwngtgyvdwlgysyngan-plyaeydvw---kyl 237

```

RESULT 4

```

AAR03775 standard; protein: 234 AA.
ID      AAR03775
XX
AC      AAR03775;
XX
DT      31-JUL-1990 (first entry)
XX
DE      Thermostable beta-glucanase.
XX
KM      Thermostable beta-glucanase; ss; Bacillus subtilis.
XX
OS      Bacillus macerans.
XX
PN      DD272102-A.
XX
PD      27-SEP-1989.
XX
PF      12-MAY-1988; 88DD-0315706.
XX
PR      12-MAY-1988; 88DD-0315706.
XX
PA      (DEAK ) AKAD WISSENSCHAFT DDR.
XX
PI      Borriess R;
XX
DR      WPI; 1990-067913/10.
XX
DR      N-PSDB; AAQ03519.
XX
PT      Thermostable beta-glucanase production -
PT      using Bacillus subtilis transformed with gene from Bacillus
PT      macerans.
XX
PS      Fig 1. ; 9pp: German.
XX
CC      The gene encoding Bacillus macerans thermostable
CC      beta-glucanase is expressed in Bacillus subtilis. The
CC      enzyme is useful for lowering the viscosity of brewing
CC      mashes and in the production of feedstuff.
XX
SQ      Sequence 234 AA;

```

Query Match 12.5%; Score 228; DB 11; Length 234;

Best Local Similarity 34.3%; Pred. No. 1.9e-11; Matches 60; Conservative 25; Mismatches 74; Indels 16; Gaps 8;

```

QY      23  TNVSAKDFSGAELTYLEVOYGFEDARKMAAASGVSSMFLYONGSEIADGRPWVEVDI 82
      72  tsasynkfdeceyrstnlgyglyevsmkpkaknlgivssftlycpr---ahgtqwdeldi 128
Db      83  EVLGKNGSPGFSNIIITGKAGAKTSEKHHAASPAADQAFHTYGLEMTNRYRWYVDDGGEV 142
      129  eflgkdtlcvfnytlngvgqh---ekvisigfaskgfhyaftwqpyikwyvdg-vl 184
QY      143  RKTGGGVSNLTGTQG-LRFNLMSEESA-AWVGOFDESKLPFQFINWVKYKXT 195
      185  kht---atanlpsrpgkimmlwngtgyvdwlgysyngan-plyaeydvw---kyl 232
Db

```

RESULT 5

```

AAR05803 standard; protein: 237 AA.
ID      AAR05803
XX
AC      AAR05803;
XX
DT      08-NOV-1990 (first entry)
XX
DE      Heat-stable endo-beta-1,3-1,4-glucanase.
XX
KM      Endo-beta-1,3-1,4-glucanase; barley; brewing.
XX
OS      Bacillus macerans.
XX
PN      DD275704-A.
XX
PD      31-JAN-1990.
XX
PF      23-SEP-1988; 88DD-0320082.
XX
PR      23-SEP-1988; 88DD-0320082.
XX
PA      (DEAK ) AKAD WISSENSCHAFT DDR.
XX
PI      Borriess R, Wobus U, Mendel R-R, Baumlein H;
XX
DR      WPI; 1990-210631/28.
XX
DR      N-PSDB; AAQ05167.
XX
PT      Prep'n. of barley plants expressing heat stable beta-glucanase -
PT      by transforming cells with appropriate vector then regeneration
PT      giving seeds useful in brewing without conversion to malt.
XX
PS      Disclosure; ; p; German.
XX
CC      The corresp. gene and the plant promoter region are inserted into an
CC      expression vector and used to transform barley cells. The transformants
CC      are used to regenerate barley plants which are useful in brewing. This
CC      protein accumulates in the ripe (but ungerminated) seeds. The
CC      proportion of these to malted seeds is increased therefore, without
CC      an unacceptable increase in viscosity.
XX
SQ      Sequence 237 AA;

```

Query Match 12.5%; Score 228; DB 11; Length 237;

Best Local Similarity 34.3%; Pred. No. 1.9e-11; Matches 60; Conservative 25; Mismatches 74; Indels 16; Gaps 8;

```

QY      23  TNVSAKDFSGAELTYLEVOYGFEDARKMAAASGVSSMFLYONGSEIADGRPWVEVDI 82
      75  tsasynkfdeceyrstnlgyglyevsmkpkaknlgivssftlycpr---ahgtqwdeldi 131
Db      83  EVLGKNGSPGFSNIIITGKAGAKTSEKHHAASPAADQAFHTYGLEMTNRYRWYVDDGGEV 142
      132  eflgkdtlcvfnytlngvgqh---ekvisigfaskgfhyaftwqpyikwyvdg-vl 187
QY      143  RKTGGGVSNLTGTQG-LRFNLMSEESA-AWVGOFDESKLPFQFINWVKYKXT 195
      188  kht---atanlpsrpgkimmlwngtgyvdwlgysyngan-plyaeydvw---kyl 235
Db

```

RESULT 6

AAR06622 standard; protein: 237 AA.

ID AAR06622;

AC AAR06622;

DT 09-JAN-1991 (first entry)

DE Hybrid (1,3-1,4)-pre-beta-glucanase.

XX Hybrid pre-beta-glucanase; glucans; beer; animal feed; poultry.

OS Bacillus amyloliquefaciens, Bacillus macerans.
 XX Key Location/Qualifiers
 FH Domain 1..127
 FT Domain /label=amino-terminal half of B.macerans beta-glucanase
 FT Domain 131..234
 FT Domain /label=carboxyl-terminal half of B.amyloliquefaciens
 FT Domain beta-glucanase
 PN MO9009436-A.
 XX 23-AUG-1990.
 PD 16-FEB-1990; 90MO-DK00044.
 PF 04-AUG-1989; 89DK-0003848.
 PR 16-FEB-1989; 89DD-0325800.
 XX (CARL-) CARLSBERG A/S.
 PA (DEAK) AKAD WISSENSCHAFT DDR.
 PI Boriss R, Hofemeister J, Thomsen KK, Olsen O, Vonwetstein D;
 XX MPI; 1990-275129/36.
 DR N-PSDB; AA005833.
 XX
 PT New thermostable (1,3-1,4)-beta-glucanase - prep'd using hybrid
 PT gene obtd. using Bacillus amyloliquefaciens and B.macerans genes
 PS Disclosure: page 28; 84pp; English.
 XX
 CC This hybrid protein is encoded by the beta-glucanase-H2 gene.
 CC Following processing of the signal peptide the mature protein
 CC is produced, comprising the amino terminus of the B.macerans
 CC beta-glucanase and the carboxyl-terminal half of the B.amylo-
 CC liquefaciens beta-glucanase. This hybrid protein is thermo-
 CC stable and hydrolyses beta-glycosidic linkages in (1,3-1,4)-
 CC beta-glucans. Reducing sugars are obtd. at high temps. and
 CC thus this enzyme can be used in the mfr. of food prods., esp.
 CC beer and animal feed (e.g. for feeding poultry). See also
 CC AA005833.
 CC
 SQ Sequence 237 AA;

Query Match 12.2%; Score 222.5; DB 11; Length 237;
 Best Local Similarity 33.1%; Pred. No. 5.5e-11;
 Matches 57; Conservative 25; Mismatches 79; Indels 11; Gaps 6;
 OY 23 TNSAKDFSGAELTYLLEVOYGFPEARMKMAASGVSSMFLYONGSEIADGRPWEVDI 82
 DB 75 tsaaykktfdcaeyrtnlygylyevsmkpkatgtlyvsfftytgp---ahgqweidl 131
 OY 83 EVLGKRPQSQSNITITGKAGAKTSEKHAHVSPAADOAFHTYGLLEWTPNVRVMTVDGQEV 142
 DB 132 eflgkktctkvgfnyttngqg---nhkftadlfgdaanayhlyafdgqpsikwyvdgq-1 187
 OY 143 RTEGGGVSNLTGTGRLRLMSSES-AMVGFDESKLPLFQFINMVVYK 193
 DB 188 khtatqldpaapq--immllmngtgvddwlgsgnvn-plyahymtyrk 236

RESULT 7
 AAM37884
 ID AAM37884 standard; Protein; 245 AA.
 AC AAM37884;
 XX
 DT 20-AUG-1998 (first entry)
 XX
 DE Lichenase protein.
 XX
 KW Lichenase; lica; fungus; enzyme; beta-1,4-glucan bond hydrolysatlon;

KM beta-1,3-linked glucan bond; grain-containing feed; grain treatment;
 KM nutrient availability; brewing; fermentation.
 XX
 OS Orpinomyces sp.
 FH Key Location/Qualifiers
 FT Peptide 1..29
 FT Protein /note="signal peptide"
 FT Protein /note="mature lichenase"
 PN MO9814595-A1.
 XX 09-APR-1998.
 PD 03-OCT-1997; 97MO-US17811.
 PF 04-OCT-1996; 96US-0027882.
 PR (UYGE-) UNIV GEORGIA RES FOUND INC.
 PA Chen H, Li X, Ljungdahl LG;
 XX MPI; 1998-240094/21.
 DR N-PSDB; AAV29067.
 XX
 PT New isolated lichenase protein - is obtained from Orpinomyces PC-2,
 PT used for treatment of grain to improve feeds or to improve brewing
 PT and fermentation processes
 PS Claim 1; Page 24-25; 41pp; English.
 XX
 CC This sequence is the lichenase (lica) of Orpinomyces sp. strain PC-2
 CC of the invention. The protein was purified from a fungus or a fungal
 CC culture or from a recombinant DNA molecule having a fungal lichenase
 CC coding sequence. The lichenase enzyme hydrolyses the beta-1,4-glucan
 CC bonds adjacent to beta-1,3-linked glucan bonds, but does not cleave
 CC beta-1,4-linked glucans. They can be used for the treatment of animal
 CC grain-containing feeds to improve nutrient availability and for treatment
 CC of grain (e.g. barley or wheat) in the brewing and fermentation
 CC industries to increase carbon substrate availability and to maximise
 CC production of desired products.
 CC
 SQ Sequence 245 AA;

Query Match 10.4%; Score 190.5; DB 19; Length 245;
 Best Local Similarity 31.6%; Pred. No. 2.8e-08;
 Matches 54; Conservative 17; Mismatches 87; Indels 13; Gaps 6;
 OY 21 LFTNVSADKDFSGAELTYLLEVOYGFPEARMKMAASGVSSMFLYONGSEIADGRPWEV 80
 DB 79 lldrdgsgyltceyrtknyygmfgvmkpklnpdyvvsfftytgp---dgtkwdei 135
 OY 81 DIEVLKNGPQSQSNITITGKAGAKTSEKHAHVSPAADOAFHTYGLLEWTPNVRVMTVDGQ 140
 DB 136 dleflgdtktvgfnyttngqgh---enhlylgfdaagfhygffwarsnstlywvdc 192
 OY 141 EVKRTGGGVSNLTGTG-RLRLMSSES-AMVGFDESKLPLFQFINMV 189
 DB 193 ayvla---ydnldtprkimmawnglgvddwlrpf- gtnlisygdwv 238

RESULT 8
 AAC48467
 ID AAC48467 standard; Protein; 282 AA.
 AC AAC48467;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 61207.
 XX

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135121.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

Accession	Gene	Protein	Protein ID	Protein Name	Protein Description
XX	17-OCT-2000	(first entry)			
XX		Arabidopsis thaliana protein fragment SEQ ID NO: 39167.			
DE		Protein identification: signal transduction pathway: metabolic pathway:			
XX		hybridisation assay; genetic mapping; gene expression control; promoter			
KM		termination sequence.			
XX		Arabidopsis thaliana.			
OS		EP1033405-A2.			
PN					
XX					
PD	06-SEP-2000.				
XX					
PF	25-FEB-2000; 2000EP-0301439.				
XX					
PR	25-FEB-1999; 99US-0121825.				
PR	05-MAR-1999; 99US-0123180.				
PR	09-MAR-1999; 99US-0123548.				
PR	23-MAR-1999; 99US-0125788.				
PR	25-MAR-1999; 99US-0126264.				
PR	29-MAR-1999; 99US-0126785.				
PR	01-APR-1999; 99US-0127462.				
PR	06-APR-1999; 99US-0128234.				
PR	08-APR-1999; 99US-0128714.				
PR	16-APR-1999; 99US-0129845.				
PR	19-APR-1999; 99US-0130077.				
PR	21-APR-1999; 99US-0130449.				
PR	23-APR-1999; 99US-0130510.				
PR	23-APR-1999; 99US-0130891.				
PR	28-APR-1999; 99US-0131449.				
PR	30-APR-1999; 99US-0132048.				
PR	30-APR-1999; 99US-0132407.				
PR	04-MAY-1999; 99US-0132484.				
PR	05-MAY-1999; 99US-0132485.				
PR	06-MAY-1999; 99US-0132486.				
PR	06-MAY-1999; 99US-0132487.				
PR	07-MAY-1999; 99US-0132863.				
PR	11-MAY-1999; 99US-0134256.				
PR	14-MAY-1999; 99US-0134218.				
PR	14-MAY-1999; 99US-0134219.				
PR	14-MAY-1999; 99US-0134221.				
PR	14-MAY-1999; 99US-0134370.				
PR	18-MAY-1999; 99US-0134768.				
PR	19-MAY-1999; 99US-0134941.				

PR	21-MAY-1999	99US-0135353.
PR	24-MAY-1999	99US-0135629.
PR	25-MAY-1999	99US-0136021.
PR	27-MAY-1999	99US-0136392.
PR	28-MAY-1999	99US-0136782.
PR	01-JUN-1999	99US-0137222.
PR	03-JUN-1999	99US-0137528.
PR	04-JUN-1999	99US-0137502.
PR	07-JUN-1999	99US-0137724.
PR	08-JUN-1999	99US-0138094.
PR	10-JUN-1999	99US-0138540.
PR	10-JUN-1999	99US-0138847.
PR	14-JUN-1999	99US-0139119.
PR	16-JUN-1999	99US-0139452.
PR	16-JUN-1999	99US-0139453.
PR	17-JUN-1999	99US-0139492.
PR	18-JUN-1999	99US-0139454.
PR	18-JUN-1999	99US-0139455.
PR	18-JUN-1999	99US-0139456.
PR	18-JUN-1999	99US-0139457.
PR	18-JUN-1999	99US-0139458.
PR	18-JUN-1999	99US-0139459.
PR	18-JUN-1999	99US-0139460.
PR	18-JUN-1999	99US-0139461.
PR	18-JUN-1999	99US-0139462.
PR	18-JUN-1999	99US-0139463.

PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156506.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.5%, Score 173; DB 21; Length 282;
Best Local Similarity 25.6%; Pred. No. 1e-06;
Matches 64; Conservative 36; Mismatches 92; Indels 58; Gaps 12;

OY 24 NVSAKDFGAEIYTLAEVOYGFKEARMKMAA--ASGTVSSMFLYONGSEIDGRPMVEVD 81
DB 53 slsldkssgsqfgsnqgefilygaewqmklvpnsqglvtlyfik-----spgltwdeid 106
OY 82 IEVLCK--NPGSFOSNITITGRGAOKTSEKHAVSPADDAFHITYGLEWTPNVRWTVD 138
DB 107 feflgnisghpytlhntlytc-kgtdkqgqflwfdpevn--fhlyctlmprgriffv 163
OY 139 GOEVRKTBGGOVSNL--TGTGCLRF--NLMSSESAAMVG--QPDSEKLPLEFQFINMKV 191
DB 164 gplrefknpaiaigyplrtgpmrlyaslweahwatrgjlektwskapftaf----- 217
OY 192 KYRTPGQGGGSDFTLMDTNDPFDGSRWCGKDWTFDGNRVLDLTKNIVYRDMILIAL 251
DB 218 yr-----nyhv-dgcvwangksacsanpwtfgk----- 246
OY 252 TRKGOESFNG 261

Db 247 dsngqtrmkq 256

RESULT 10
AAG52114
ID AAG52114 standard; Protein: 277 AA.
XX
AC AAG52114;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 66212.
XX
KM Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay: genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 09-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 22-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136782.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142890.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145818.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

```
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155655.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160881.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.5%; Score 172.5; DB 21; Length 277;
Best Local Similarity 26.6%; Pred. No. 1, le-06;
Matches 54; Conservative 34; Mismatches 74; Indels 41; Gaps 9;
```

```
OY 24 NVSADFGALYLTLEVOYGFEMRMKAA--ASGVSMFLYNGSEIADGRWEVD 81
Db 48 slsldkssgsgfsgqeflygkaevgmklvpqnsagtyltfyllk-----spgltwdeid 101
OY 82 IEVLCK---NGSGFSQSNITITGAKAGAKTSEKHAVSPAADAFHHTYGLGEMPNYVRWTVVD 138
Db 102 feflgnlsgnhytlnhnyvt-kgsqdkqgfh\wfdptan--fhnycltmprqrliflvd 158
```

```
OY 139 GQEVKRTREGQVSNLTGTQGLNF-----NLMSSESAAMWG--QFDESKLPIRF 185
Db 159 gplrefmame-----grgyfpkqgmrllyaslweehwatrglktktskaptfay 212
OY 186 IN-----WKVKRYTPGCGE 200
Db 213 yrnynvegvcwvngksvcpsansq 235

RESULT 11
AAG18644
ID AAG18644 standard; Protein; 269 AA.
AC AAG18644;
XX
XX 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 20135.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX
XX EP103405-A2.
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134371.
PR 18-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 19-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
```

PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139452.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140891.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 02-JUL-1999; 99US-0142154.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142820.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 28-JUL-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151338.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160880.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 9.3%; Score 170.5; DB 21; Length 269;
 Best Local Similarity 27.6%; Pred. No. 1.5e-06;
 Matches 59; Conservative 35; Mismatches 97; Indels 23; Gaps 10;

OY 24 NSAKDFSCAELTYTLEVOYKFEAPKMAA--ASGTSSMFLYONGSEIADCRPWEVD 81
 DB 40 slsldkssgsgfsgnqeflygkaevgmklvpgnsagvtltfYlk-----spgtwdeid 93

PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149375.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149920.
PR 23-AUG-1999; 9905-0149920.
PR 23-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 30-AUG-1999; 9905-0151080.
PR 31-AUG-1999; 9905-0151303.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 28-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161922.
PR 28-OCT-1999; 9905-0161933.
PR 29-OCT-1999; 9905-0162142.

Query Match 9.38; Score 170.5; DB 21; Length 282;
Best Local Similarity 27.6%; Pred. No. 1.6e-06;

Matches 59; Conservative 35; Mismatches 97; Indels 23; Gaps 10;
QY 24 NVSAKDFSGALVLYTEVQKFEARKMMAA--ASTVSSWFLYQNSSEIADGRPWTEVD 81
Db 53 SLSDKSSGSGfGNGefLYKkaevqmkLVpGnsagvtLfyL-----spgtLweid 106
QY 82 IEVLGK--NPGSPQSNIIITGKAGAQKTSKHNHNVSPAADQAFHTYGLFWTPNVVMTVD 138
Db 107 feflgnlsgnphylthrvyc-ksgdqkeqfhlwfdptvn--fhtyeltmnpqrlftvd 163
QY 139 GQEVKRTGCGVSNL--TGQGLRF--NLMSSESAWV--QFDESKLPLFOFINMKV 191
Db 164 gipLrefkneslgyvpfptkGpmfLyaslweaehwatrglJektwskapftafyrnyv 223
QY 192 YKYPGGGEGGSDFTLDM--TDNFDTPDGRMKG 224
Db 224 egcwangksccpansswfctqldos-ngqlcmky 256
RESULT 13
AAAG18642
ID AAAG18642 standard; Protein; 298 AA.
XX AC AAAG18642;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 20133.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EPI1033405-A2.
XX PD 06-SEP-2000.
XX PE 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 9905-0121825.
XX PR 05-MAR-1999; 9905-0123180.
XX PR 09-MAR-1999; 9905-0123548.
XX PR 23-MAR-1999; 9905-0125788.
XX PR 25-MAR-1999; 9905-0126264.
XX PR 29-MAR-1999; 9905-0126785.
XX PR 01-APR-1999; 9905-0127462.
XX PR 06-APR-1999; 9905-0128234.
XX PR 08-APR-1999; 9905-0128714.
XX PR 16-APR-1999; 9905-0129845.
XX PR 19-APR-1999; 9905-0130077.
XX PR 21-APR-1999; 9905-0130049.
XX PR 23-APR-1999; 9905-0130510.
XX PR 23-APR-1999; 9905-0130891.
XX PR 28-APR-1999; 9905-0131449.
XX PR 30-APR-1999; 9905-0132048.
XX PR 04-MAY-1999; 9905-0132407.
XX PR 04-MAY-1999; 9905-0132484.
XX PR 05-MAY-1999; 9905-0132485.
XX PR 06-MAY-1999; 9905-0132486.
XX PR 06-MAY-1999; 9905-0132487.
XX PR 07-MAY-1999; 9905-0132863.
XX PR 11-MAY-1999; 9905-0134256.
XX PR 14-MAY-1999; 9905-0134218.
XX PR 14-MAY-1999; 9905-0134219.
XX PR 14-MAY-1999; 9905-0134221.
XX PR 14-MAY-1999; 9905-0134370.
XX PR 18-MAY-1999; 9905-0134768.
XX PR 19-MAY-1999; 9905-0134941.
XX PR 20-MAY-1999; 9905-0135124.
XX PR 21-MAY-1999; 9905-0135353.
XX PR 24-MAY-1999; 9905-0135629.

PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	14-JUN-1999;	99US-0138847.
PR	16-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142803.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142970.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145086.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	16-AUG-1999;	99US-0148684.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136332.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142290.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145513.
PR 27-JUL-1999; 99US-0145518.
PR 27-JUL-1999; 99US-0145519.
PR 28-JUL-1999; 99US-0145591.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161358.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 8.6%; Score 157.5; DB 21; Length 269;
 Best Local Similarity 25.2%; Pred. No. 1.9e-05;
 Matches 56; Conservative 39; Mismatches 76; Indels 51; Gaps 10;

OY 6 TAVKSAALVAAAAALTTNVS-----AKDFSGAEIYTL-----EVOYGFKE 47
 Db 11 tllvaatsvsadfnldvnavgnggrkllnngqlltllsldksqsgfqskeylelfgkld 70
 OY .48 ARMKAA--ASGTSSMFLYONGSEIADGRPMVEVDIEVLCK--NPGSFOSNITGKAG 102
 Db 71 mqiklvpgnsagvlctfkyiksegs-----twdeiddeflgnmsgdpylhcnvycgqkg 124
 OY 103 AOKTSEKHAAVSPADQAFHTYGLFMTPNYVMYVDSGOEVRKTEGGOVSNLTGTGLRF- 161
 Db 125 -dkeqgfhlwfdptan--fhysilwmpqrllltvddtplefknqe-----slgvllp 175
 OY 162 -----NMWSESAAWVG---OFDESKLPLFOFINMKV 191
 Db 176 knkpmrmyasilwnaddwatrglvtgtdwskapfmasyrniki 217

Search completed: August 23, 2001, 11:14:29
 Job time: 56 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:18:02 ; Search time 32.82 seconds

(without alignments)
619,702 Million bell updates/sec

Title: us-09-654-652a-2

Perfect score: 267

Sequence: 1 MVSADFGSAELTYLLEVOY.....PNSSSVDKLAALLENHHNNH 267

Scoring table: OLIGO

Searched: 219241 seqs, 76174552 residues

Word size: 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	247	92.5	349	2 A44507	licheninase (EC 3.
2	8	3.0	381	2 S29560	fructose-bisphosph
3	8	3.0	407	2 T06408	probable fructose-
4	8	3.0	429	2 T07815	S-locus-specific g
5	8	3.0	429	2 D70784	probable Rieske [2
6	7	2.6	60	2 C64698	probable histidine
7	7	2.6	95	2 S55439	hypothetical prote
8	7	2.6	115	2 P84469	probable glycine-r
9	7	2.6	115	2 H72583	hypothetical prote
10	7	2.6	119	2 S64079	probable membrane
11	7	2.6	139	2 T41242	hypothetical lysin
12	7	2.6	139	2 T41526	hypothetical prote
13	7	2.6	206	2 T25384	hypothetical prote
14	7	2.6	259	2 A70359	hydrogenase expres
15	7	2.6	270	2 A26480	knob protein - mal
16	7	2.6	304	2 F69864	hypothetical prote
17	7	2.6	306	2 A83551	still frameshift p
18	7	2.6	317	2 F82672	Arp sulfurylase, s
19	7	2.6	348	2 T08814	1,3-beta-glucanase
20	7	2.6	351	1 KG20HL	histidine-rich gly
21	7	2.6	377	2 A32548	homeobox protein M
22	7	2.6	396	1 S15992	flavonemoglobin hm
23	7	2.6	396	2 F85900	hypothetical prote
24	7	2.6	420	2 T36193	probable salicylat
25	7	2.6	440	2 D64186	conserved hypotet
26	7	2.6	468	1 TVHUSE	transcription fact
27	7	2.6	469	1 TVHUSE	transcription fact
28	7	2.6	470	2 T20851	hypothetical prote
29	7	2.6	473	2 A54494	knob-associated hi

30	7	2.6	474	2 S75464	hypothetical prote
31	7	2.6	476	2 S57963	methyl Cpg binding
32	7	2.6	492	2 A41907	transcription fact
33	7	2.6	495	1 S31223	hypothetical prote
34	7	2.6	510	2 T20850	beta-galactosidase
35	7	2.6	546	2 B32688	conserved hypotet
36	7	2.6	611	2 F70325	knob protein precu
37	7	2.6	634	2 A54495	histidine-rich pro
38	7	2.6	637	2 A28412	probable fadD32 pr
39	7	2.6	637	2 E70887	knob-associated hi
40	7	2.6	654	2 B71623	knob-associated hi
41	7	2.6	657	2 A29454	lamin A - African
42	7	2.6	665	2 S02358	beta-galactosidase
43	7	2.6	677	2 A32611	hypothetical prote
44	7	2.6	715	2 H84799	hypothetical prote
45	7	2.6	955	2 T48515	hypothetical prote

ALIGNMENTS

RESULT 1
A44507
licheninase (EC 3.2.1.73) - Fibrobacter succinogenes
C:Species: Fibrobacter succinogenes
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 08-Oct-1999
C:Accession: A44507
R:Teather, R.M.; Erfle, J.D.
J. Bacteriol. 172, 3837-3841, 1990
A:Title: DNA sequence of a Fibrobacter succinogenes mixed-linkage beta-glucanase (1.3
A:Reference number: A44507; MUID:90299807
A:Accession: A44507
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <TEA>
A:Cross-references: EMBL:M33676; NID:G148575; PIDN:AAA24896.1; PID:G148576
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 92.5%; Score 247; DB 2; Length 349;
Best Local Similarity 100.0%; Pred. No. 1.9e-244;
Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 VSADFGSAELTYLLEVOYGFKEARMKMAASGVSSMFLYONGSEIADGRPMVEVDLEV 61
DB 25 VSADFGSAELTYLLEVOYGFKEARMKMAASGVSSMFLYONGSEIADGRPMVEVDLEV 84
QY 62 LKNPGEFOSNITITGKAGAKTSEKHAHVSPADQAFHTYGLWTPNVVMTVDQEVK 121
DB 85 LKNPGEFOSNITITGKAGAKTSEKHAHVSPADQAFHTYGLWTPNVVMTVDQEVK 144
QY 122 TEGGVSNLTGTGCLRFNLMSSSAAMVGPDESKLPLFQFNNVKKYKYPGGGEGSSD 181
DB 145 TEGGVSNLTGTGCLRFNLMSSSAAMVGPDESKLPLFQFNNVKKYKYPGGGEGSSD 204
QY 182 FTLDFTDNFDPDGRMGKGMTPFGNVRDLTDKNYSRDGLILALRKQGESFNGQVP 241
DB 205 FTLDFTDNFDPDGRMGKGMTPFGNVRDLTDKNYSRDGLILALRKQGESFNGQVP 264
QY 242 RDDEPAP 248
DB 265 RDDEPAP 271
RESULT 2
S29560
fructose-bisphosphatase (EC 3.1.3.11) - garden pea (fragment)
C:Species: Pisum sativum (garden pea)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C:Accession: S29560
R:Carasco, J.L.; Chueca, A.; Hermoso, R.; Lazaro, J.J.; Ramos, J.L.; Sahrawy, M.; Pr
submitted to the EMBL Data Library, October 1992
A:Description: Cloning, structure and expression of a pea cDNA clone coding for a pho

A:Reference number: 529560
A:Accession: 529560
A:Molecule type: mRNA
A:Residues: 1-381 <CAR>
A:Cross-references: EMBL:X68826; NID:g20716; PIDN:CAA48719.1; PID:g20717
C:Superfamily: fructose-bisphosphatase
C:Keywords: phosphoric monoester hydrolase

Query Match 3.0%; Score 8; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 128 SMLGTG 135
|||||||
DB 86 SMLGTG 93

RESULT 3
T06408
Probable fructose-bisphosphatase (EC 3.1.3.11) precursor - garden pea chloroplast
C:Species: chloroplast *Plum sativum* (garden pea)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Jun-1999
C:Accession: T06408
R:Hahn, T.R.; Dong, S.M.; Rhim, J.H.
Submitted to the EMBL Data Library, February 1999
A:Description: cDNA sequence and red light modulated expression of chloroplast fructose-
A:Reference number: Z15655
A:Accession: T06408
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-407 <IAH>
A:Cross-references: EMBL:L24806; NID:g609560; PIDN:AD10213.1; PID:g609561
A:Experimental source: cv. Giant
C:Genetics:
A:Genome: chloroplast
C:Superfamily: fructose-bisphosphatase
C:Keywords: chloroplast; phosphoric monoester hydrolase
F:1-50/Domain: transit peptide (chloroplast) #status predicted <TNP>
F:51-407/Product: fructose-bisphosphatase #status predicted <MAN>

Query Match 3.0%; Score 8; DB 2; Length 407;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 128 SMLGTG 135
|||||||
DB 112 SMLGTG 119

RESULT 4
T07815
S-locus-specific glycoprotein S8 - radish (fragment)
C:Species: *Raphanus sativus* (radish)
C>Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
C:Accession: T07815
R:Sakamoto, K.; Kusaba, M.; Nishio, T.
Mol. Gen. Genet. 258, 397-403, 1998
A:Title: Polymorphism of the S-locus glycoprotein gene (SLG) and the S-locus related gen
A:Reference number: Z16146; MUID:98311079
A:Accession: T07815
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-429 <SAK>
A:Cross-references: EMBL:AB009683; NID:g3327851; PIDN:BA431730.1; PID:g3327852
C:Genetics:
A:Gene: SLG(S8)
C:Superfamily: S-locus-specific glycoprotein; S-locus-specific glycoprotein homology
C:Keywords: glycoprotein
F:32-427/Domain: S-locus-specific glycoprotein homology <SSG>

Query Match 3.0%; Score 8; DB 2; Length 429;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 213 TDKNIYSR 220
|||||||
DB 253 TDKNIYSR 260

RESULT 5
D70784
Probable Rieske [2Fe-2S] iron-sulfur protein cyoB - Mycobacterium tuberculosis (strai
C:Species: *Mycobacterium tuberculosis*
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: D70784
R:Coler, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtrold,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: D70784
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-429 <COL>
A:Cross-references: GB:Z70283; GB:AL123456; NID:g3261561; PIDN:CAA94264.1; PID:e23357
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: qcrA
C:Superfamily: Rieske [2Fe-2S] homology
C:Keywords: 2Fe-2S; metalloprotein; Rieske iron-sulfur protein
F:343-389/Domain: Rieske [2Fe-2S] homology <RSK>
F:353,355,372,375/Binding site: 2Fe-2S cluster (Cys, His, Cys, His) (covalent) #statu

Query Match 3.0%; Score 8; DB 2; Length 429;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 230 RKGOESFN 237
|||||||
DB 335 RKGOESFN 342

RESULT 6
C64698
Probable histidine-rich metal-binding protein - *Helicobacter pylori*
C:Species: *Helicobacter pylori*
A:Variety: strains J99, 26695
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: C64698; C71821
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khajak, H.G.; Glodek, A.; McKe
non, J.D.; Kelley, J.M.; Cotton, M.D.; Feldman, J.M.; Fujii, C.; Bowman, C.; Watney,
Nature 388, 539-547, 1997
A:Authors: Mallin, E.; Hayes, W.S.; Borodovsky, M.; Karik, P.D.; Smith, H.O.; Fraser,
A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
A:Reference number: A64520; MUID:9794467
A:Accession: C64698
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-60 <TOM>
A:Cross-references: GB:AE000643; GB:AE000511; NID:g2314598; PIDN:AA008471.1; PID:g231
A:Experimental source: strain 26695
R:Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, O.; Taylor, D.E.; Vovis, G.F.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
A:Reference number: A71800; MUID:99120557
A:Accession: C71821
A:Molecule type: DNA
A:Residues: 1-60 <ARN>

A:Cross-references: GB:AE001555; GB:AE001439; NID:g4155929; PIDN:AAD06898.1; PID:g415593
 A:Experimental source: strain j99
 C:Genetics:
 A:Gene: HP1427; jhp1320

Query Match 2.6%; Score 7; DB 2; Length 60;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 261 EHHNNH 267
 DB 27 EHHNNH 33

RESULT 7
 S55439
 hypothetical protein ywkr - Bacillus subtilis

C:Species: Bacillus subtilis
 C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 15-Oct-1999
 C:Accession: S55439; E70061
 R:Glaser, P.; Danchin, A.
 submitted to the EMBL Data Library, May 1995
 A:Description: Cloning and sequencing of the Bacillus subtilis chromosomal region from 3
 A:Reference number: S55414
 A:Accession: S55439
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-95 <GLA>
 A:Cross-references: EMBL:Z49782; NID:9853752; PIDN:CA69886.1; PID:9853778
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chc
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
 lech, J.; Harwood, C.R.; Henuat, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.
 R.; Koenigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033
 A:Accession: E70061
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-95 <KUN>
 A:Cross-references: GB:Z99122; GB:AL009126; NID:g2636029; PIDN:CA615716.1; PID:e1184605;
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: ywkr

Query Match 2.6%; Score 7; DB 2; Length 95;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 RKKMAA 32
 DB 71 RKKMAA 77

RESULT 8
 F44469

probable glycine-rich protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: F44469
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanKen, S.E.; Umayam, L.; Tallon, L.

euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A64420; MUID:20083487
 A:Accession: F44469

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-115 <STO>
 A:Cross-references: GB:AE002093; NID:g4581171; PIDN:AAD24654.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g05530
 A:Map position: 2

Query Match 2.6%; Score 7; DB 2; Length 115;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 AAAGTV 36
 DB 22 AAAGTV 28

RESULT 9
 H72583
 hypothetical protein APE1948 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: H72583
 R:Kavayatsyl, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Ta
 ana, H.; Takamiza, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
 DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero

A:Reference number: A72450; MUID:99310339
 A:Accession: H72583
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-115 <KAW>
 A:Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAA80957.1; PID:dl044743; PID:g
 C:Genetics:
 A:Experimental source: strain K1
 A:Gene: APE1948

Query Match 2.6%; Score 7; DB 2; Length 115;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 261 EHHNNH 267
 DB 31 EHHNNH 37

RESULT 10
 S64079

probable membrane protein YGL072c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein G3263
 C:Species: Saccharomyces cerevisiae

C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 09-Jun-2000
 C:Accession: S64079
 R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
 submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64071
 A:Accession: S64079
 A:Molecule type: DNA
 A:Residues: 1-119 <RID>
 A:Cross-references: EMBL:Z72596; NID:g1322584; PID:e243291; PID:g1322587; GSPDB:GN000
 C:Genetics:
 A:Gene: MIPS:YGL072c
 A:Map position: 7L
 C:Superfamily: Saccharomyces cerevisiae probable membrane protein YGL072c
 C:Keywords: transmembrane protein

F:35-51/Domain: transmembrane #status predicted <TM>

Query Match 2.6%; Score 7; DB 2; Length 119;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 FSGAELY 13
|111111
DB 98 FSGAELY 104

RESULT 11

T41242

hypothetical lysine-rich protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T41242

R:Lyne, M.; Brown, S.; Quail, M.; Harris, D.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, July 1999

A:Reference number: 221980

A:Accession: T41242

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-139 <LYN>

A:Cross-references: EMBL:AL109608; PIDN:CAB51561.1; GSPDB:GN00068; SPDB:SPCC23B6.02C

A:Experimental source: strain 972h-; cosmid c23B6

A:Gene: SPDB:SPCC23B6.02C

A:Map position: 3

A:Introns: 106/1

Query Match 2.6%; Score 7; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 DKLAAL 260
|111111
DB 74 DKLAAL 80

RESULT 12

T41526

hypothetical protein SPCC645.09 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T41526

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, March 1999

A:Reference number: 222000

A:Accession: T41526

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-139 <MOO>

A:Cross-references: EMBL:AL049498; PIDN:CAB39905.1; GSPDB:GN00068; SPDB:SPCC645.09

A:Experimental source: strain 972h-; cosmid c645

C:Genetics:

A:Gene: SPDB:SPCC645.09

A:Map position: 3

Query Match 2.6%; Score 7; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 DEPAFNS 250
|111111
DB 112 DEPAFNS 118

RESULT 13

T25384

hypothetical protein T27F6.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T25384

R:Dobson, R.
submitted to the EMBL Data Library, November 1996

A:Reference number: 220026

A:Accession: T25384

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-206 <WTL>

A:Cross-references: EMBL:Z62060; PIDN:CAB04883.1; GSPDB:GN00019; CESP:T27F6.4

A:Experimental source: clone T27F6

C:Genetics:

A:Gene: CESP:T27F6.4

A:Map position: 1

A:Introns: 61/3; 154/3

Query Match 2.6%; Score 7; DB 2; Length 206;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 EHHHHH 267
|111111
DB 115 EHHHHH 121

RESULT 14

A70359

hydrogenase expression/formation protein B - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 24-Sep-1999

C:Accession: A70359

R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666

A:Accession: A70359

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-259 <AQF>

A:Cross-references: GB:AE000701; NID:g2983260; PIDN:AAC06860.1; PID:g2983266; GB:AE00

C:Experimental source: strain VF5

C:Genetics:

A:Gene: hypB

C:Superfamily: hydrogenase expression/formation protein hypB

Query Match 2.6%; Score 7; DB 2; Length 259;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 EHHHHH 267
|111111
DB 17 EHHHHH 23

RESULT 15

A26480

knob protein - malaria parasite (Plasmodium falciparum) (fragments)
C:Species: Plasmodium falciparum
C:Date: 23-Aug-1987 #sequence_revision 23-Aug-1987 #text_change 09-Jun-2000

C:Accession: A26480

R:Kilejian, A.; Sharma, Y.D.; Karoui, H.; Naslund, L.
Proc. Natl. Acad. Sci. U.S.A. 83, 7938-7941, 1986

A:Title: Histidine-rich domain of the knob protein of the human malaria parasite Plas

A:Reference number: A26480; MUID:87017062

A:Accession: A26480

A:Molecule type: mRNA

A:Residues: 1-270 <KIL>

A:Cross-references: GB:M14210; NID:g160361; PIDN:AAA29631.1; PID:g160362

C:Superfamily: knob-associated histidine-rich protein

Query Match 2.6%; Score 7; DB 2; Length 270;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 261 EHHHHH 267
|||||
Db 62 EHHHHH 68

Search completed: August 23, 2001, 11:18:03
Job time: 210 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:15:07; Search time 32.87 Seconds

(without alignments)
618.759 Million cell updates/sec

Title: US-09-654-652a-2

Perfect score: 1439
Sequence: 1 MSASKDFSGAELTYLEVOY.....PNSSVDKLAALAEHHHHH 267

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1340	93.1	349	2	A44507 licheninase (EC 3.1.4)
2	266.5	18.5	851	2	H84053 licheninase (EC 3.1.4)
3	265.5	18.5	252	2	A48378 licheninase (EC 3.1.4)
4	247.5	17.1	242	1	LXBS licheninase (EC 3.1.4)
5	245.5	16.9	802	2	A36910 licheninase (EC 3.1.4)
6	242.5	16.6	239	1	A29091 licheninase (EC 3.1.4)
7	238.5	16.5	234	1	S23498 licheninase (EC 3.1.4)
8	237	16.5	243	1	S15388 licheninase (EC 3.1.4)
9	229.5	15.9	237	1	S11927 licheninase (EC 3.1.4)
10	226	15.7	238	1	S19012 licheninase (EC 3.1.4)
11	225	15.6	242	2	J50611 licheninase (EC 3.1.4)
12	222.5	15.5	302	2	G84053 licheninase (EC 3.1.4)
13	190.5	13.2	507	2	S64507 licheninase (EC 3.1.4)
14	182.5	12.7	268	2	S34804 licheninase (EC 3.1.4)
15	180	12.5	282	2	T02354 licheninase (EC 3.1.4)
16	172	12.0	277	2	B85354 licheninase (EC 3.1.4)
17	171.5	11.9	282	2	A85354 licheninase (EC 3.1.4)
18	169.5	11.8	642	2	B72428 licheninase (EC 3.1.4)
19	169	11.7	286	2	S71225 licheninase (EC 3.1.4)
20	163	11.3	287	2	T04236 licheninase (EC 3.1.4)
21	159	11.0	277	2	S71222 licheninase (EC 3.1.4)
22	158.5	10.7	269	2	S61555 licheninase (EC 3.1.4)
23	150.5	10.5	467	2	S30839 licheninase (EC 3.1.4)
24	147	10.2	284	2	T52097 licheninase (EC 3.1.4)
25	146.5	10.2	422	2	S48564 licheninase (EC 3.1.4)
26	142	9.9	289	2	T06166 licheninase (EC 3.1.4)
27	141.5	9.8	280	2	T02090 licheninase (EC 3.1.4)
28	140.5	9.8	310	2	A86239 licheninase (EC 3.1.4)
29	140.5	9.8	310	2	A86239 licheninase (EC 3.1.4)

30	140	9.7	305	2	G84568 probable xyloglucanase (EC 3.1.4)
31	139.5	9.7	286	2	T06202 xyloglucanase (EC 3.1.4)
32	133.5	9.3	286	2	S48201 xyloglucanase (EC 3.1.4)
33	132	9.2	295	2	T10210 xyloglucanase (EC 3.1.4)
34	131.5	9.1	292	2	T06201 xyloglucanase (EC 3.1.4)
35	129.5	9.0	1324	2	T18265 xyloglucanase (EC 3.1.4)
36	128	8.9	283	2	T07678 xyloglucanase (EC 3.1.4)
37	127.5	8.9	269	2	T05895 xyloglucanase (EC 3.1.4)
38	127	8.8	292	2	T04514 xyloglucanase (EC 3.1.4)
39	125.5	8.7	277	2	JE0156 xyloglucanase (EC 3.1.4)
40	124	8.6	299	2	D84519 xyloglucanase (EC 3.1.4)
41	123	8.5	296	2	C49539 xyloglucanase (EC 3.1.4)
42	121.5	8.4	292	2	T10211 xyloglucanase (EC 3.1.4)
43	121	8.4	284	2	T06200 xyloglucanase (EC 3.1.4)
44	121	8.4	293	2	T10523 xyloglucanase (EC 3.1.4)
45	120	8.3	877	2	JN0772 xyloglucanase (EC 3.1.4)

ALIGNMENTS

RESULT 1
A44507 licheninase (EC 3.2.1.73) - Fibrobacter succinogenes
C:Species: Fibrobacter succinogenes
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 08-Oct-1999
C:Accession: A44507
R:Teather, R.M.; Erfle, J.D.
J. Bacteriol. 172, 3837-3841, 1990
A:Title: DNA sequence of a Fibrobacter succinogenes mixed-linkage beta-glucanase (1,3
A:Reference number: A44507; MUID:90299807
A:Accession: A44507
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <TEA>
A:Cross-References: EMBL:M3676; NID:g148575; PIDN:AAA24896.1; PID:g148576
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 93.1%; Score 1340; DB 2; Length 349;
Best Local Similarity 99.6%; Pred. No. 7.8e-105;
Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VSAKDFSGAELTYLEVOYGFARFKMAASGVSSMFLYONGSEIADGRPVVEDIEV 61
DB 25 VSAKDFSGAELTYLEVOYGFARFKMAASGVSSMFLYONGSEIADGRPVVEDIEV 84
QY 62 LKKNPQSFQSNITITGKAGAKTSEKHNHVAVSPADQAFHTTGLWTPNVRWTVGQEVK 121
DB 85 LKKNPQSFQSNITITGKAGAKTSEKHNHVAVSPADQAFHTTGLWTPNVRWTVGQEVK 144
QY 122 TEGGVSNLTGQGLRFNLMSSSAWVGOFDESKLPLOFNMVYVYVTGGGEGSSD 181
DB 145 TEGGVSNLTGQGLRFNLMSSSAWVGOFDESKLPLOFNMVYVYVTGGGEGSSD 204
QY 182 FTLDWTDNFDFTGDSRWGKDMTFDGNRYDLTDKNITSRDGMILALTRKQGESFNGQVP 241
DB 205 FTLDWTDNFDFTGDSRWGKDMTFDGNRYDLTDKNITSRDGMILALTRKQGESFNGQVP 264
QY 242 RDDEPAPNSS 252
DB 265 RDDEPAPNSS 275

RESULT 2
H84053 endo-beta-1,3-1,4 glucanase (licheninase) bglis [imported] - Bacillus halodurans (stra
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 01-Dec-2000
C:Accession: H84053
R: Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Molecule type: DNA
A:Residues: 1-242 <KUN>
A:Cross-references: GB:299124; GB:AL009126; NID:g2636442; PIDN:CA815943.1; PID:g2636453
R:Experimental source: strain 168
R:Murphy, N.; McConnell, D.J.; Cantwell, B.A.
Nucleic Acids Res. 12, 5355-5367, 1984
A>Title: The DNA sequence of the gene and genetic control sites for the excreted B. subtilis reference number: A93526; MUID:84272222
A:Accession number: A22914
A:Molecule type: DNA
A:Residues: 1-203,'L',205-242 <MUR>
A:Cross-references: EMBL:X00754; NID:g39818; PIDN:CAA25328.1; PID:g685236
R:Experimental source: strain C120
A>Note: the authors translated the codon CAA for residue 29 as Lys and CCA for residue 8
R:Tezuka, H.; Yunkai, T.; Yabuuchi, S.
Agric. Biol. Chem. 53, 2335-2339, 1989
A>Title: Construction of a beta-glucanase hyperproducing *Bacillus subtilis* using the old
A:Reference number: A90026
A:Accession number: A90026
A:Molecule type: DNA
A:Residues: 1-23,'S',25-82,'S',84-242 <TEZ>
A:Cross-references: DBJ:D00518; NID:g216243; PIDN:BAA00405.1; PID:g216244
R:Experimental source: strain Y-25, clone pLE100
R:Yunkai, T.; Tezuka, H.; Yabuuchi, S.
Agric. Biol. Chem. 53, 2341-2346, 1989
A>Title: Purification and some properties of two enzymes from a beta-glucanase hyperprod
A:Reference number: A90027
A:Contents: annotation
A>Note: source was hyperproducing strain HL-25 with gene from strain Y-25
A>Note: the amino ends of the mature forms of E-1 and E-2 are pyroglutamic acid and glut
C:Genetics:
A:Gene: bglIS
C:Function:
A>Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in lichenan and c
C:Superfamily: licheninase
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation; F
F:1-28/Domin: signal sequence #status predicted <SIG>
F:29-24/Product: lichenanase #status predicted <MAT>
F:29/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) (partial) #status

Query Match 17.1% Score 245.5; DB 1; Length 242;
Best Local Similarity 35.8%; Pred. No. 2.6e-13;
Matches 59; Conservative 24; Mismatches 71; Indels 11; Gaps 6;

OY 7 FSGALTYLTLEEYOXGKFAPRMKAASGTVSSMELYONGSEIADGRPVWEVDIEVLGNP 66
| : : : | : | | | | : | | | | : | : | | | | :
Db 87 FDCCGNRSRVQRYGYGLVEVRKKPAKNIGIVSSFITYPT---DCTPWDEIDIELGNDT 143

OY 67 GSFGSNITTKAGAKOKTISEKHNAHVSPPADQAFTHTGLEMTFNVMVTVDGGEVRKEGGQ 126
| : : : | : | | | | : | : | | | | : | : | | | | :
Db 144 TKVOFNYYTNAG--NNEKITVDLGFDANAHNYAFADWDPNSIKIWYDQG-LKHTATNQ 199

OY 127 VSNLTGTGRLRFNLMSSEA-AWVGQPFDSKLPLQLFINMKVKV 170
| : : : | : | | | | : | : | | | | : | : | | | | :
Db 200 IPTTPGK--IMNMILNGTGVDMLGSYNGVN-PLYAHDWWRKYTK 241

RESULT 6
A36910
xylinase, beta(1,3-1,4)-glucanase - Ruminococcus flavefaciens
C:Species: Ruminococcus flavefaciens
C>Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
R:Accession: A36910
R:Flint, H.J.; Martlin, J.; McPherson, C.A.; Daniel, A.S.; Zhang, J.X.
J. Bacteriol. 175, 2943-2951, 1993
A>Title: A bifunctional enzyme, with separate xylinase and beta(1,3-1,4)-glucanase domat
A:Reference number: A36910; MUID:93259938
A:Accession: A36910
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-802 <FLI>
A:Cross-references: GB:S61204; NID:g385910; PIDN:AAB26620.1; PID:g385911

A:Note: sequence extracted from NCBI backbone (NCBITN:131871, NCBIPI:131872)
C:Superfamily: endo-1,4-beta-xylanase homology; Thermotoga xylanase A amino-terminal F:42-235/Domain: endo-1,4-beta-xylanase homology <YLL>
F:259-401/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA>

```
Query Match          16.9%  Score 242.5;  DB 2;  Length 802;
Best Local Similarity 34.9%  Pred. No. 2,1e-12;
Matches 58;  Conservative 21;  Mismatches 74;  Indels 13;  Gaps 6;

OY      7  FSGAELTYLEVOYQGFKEFARKMAAASGTVSMFLYONGSEIADGRPVEIDVGLKNP 66
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      634  YSGGEFRNNRHHGYICSMQAMKNDGVSSFFTYGPR---DDNPWDEIDIELGKNT 690

OY      67  GSPFSNITTKGAKGAKTSEKHHAVSPAADAQFHTYGLEMTNRYKRVTDGQVRKTEGGQ 126
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      691  TVQVFNYYTNGCGKH---EKLYDLGFDSSSEAYHLYGFDMQPNYIAMYVGDREYVRA--- 743

OY      127  VSNLTGTGQ-LRFNLMSSSA-AWVGQFDESKLPFOFINNVKYK 170
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      744  TQDIPKTEGKIMNMAWPGLTVDWMLKAFN-GRPTLTAHYQWVYNNK 788

RESULT      7
A:29091
N:licheninase (EC 3.2.1.73) beta - Bacillus amylioliquefaciens
C:Species: Bacillus amylioliquefaciens
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A29091
R:Hotemelster, J.; Kurtz, A.; Borriass, R.; Knowles, J.
Gene 49, 177-187, 1986
A:Title: The beta-glucanase gene from Bacillus amylioliquefaciens shows extensive homo
A:Reference number: A91564; MUID:87192007
A:Accession: A29091
A:Molecule type: DNA
A:Residues: 1,239 <HOF>
A:Cross-references: GB:M15674; NID:g143009; PIDN:AAA87323.1; PID:g143010
A:Experimental source: strain BE20/78
C:Genetics:
A:Gene: bglA
C:Superfamily: Licheninase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match          16.6%  Score 238.5;  DB 1;  Length 239;
Best Local Similarity 34.3%  Pred. No. 9,8e-13;
Matches 58;  Conservative 26;  Mismatches 74;  Indels 11;  Gaps 6;

OY      3  SAKDFSGAELTYLEVOYQGFKEFARKMAAASGTVSMFLYONGSEIADGRPWEVDIVL 62
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      80  SYNKFDGCGENSVQYGYGLVEYRMRPAKNGIVSSFFTYGPR---GTPWDEIDIEFL 136

OY      63  GKNPSSPFSNITTKGAKGAKTSEKHHAVSPAADAQFHTYGLEMTNRYKRVTDGQEVAKT 122
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      137  GKDJTKVQFNYYTNGAG--NHEKPADLGFDPAANAYHYAARDWOPNSITKMYVDQ-LKRT 192

OY      123  EGGGVSNLTGTGQGLRFNLMSSSA-AWVGQFDESKLPFOFINNVKYK 170
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      193  ATTQIPAAAPGK--IMMNLNMGTVDDMLGSLNGVN-PIYAHYDMWRKYK 238

RESULT      8
A:523498
N:licheninase (EC 3.2.1.73) l1cB precursor - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S23498; S22137
R:Schlimming, S.; Schwarz, W.H.; Staudenbauer, W.L.
Eur. J. Biochem. 204, 13-19, 1992
A:Title: Structure of the Clostridium thermocellum gene l1cB and the encoded beta-1,3
1 cellulases.
```



```

Db      87  MTLAKRRYDNPGLKSNFYIMCKLEVIILKANGTIVSSFYLSDDLD-----EIDIE 138
QY      61  VLGNPQSFQSNITGKAGAKTSEKHHAVSPADQAFHTYGLKMTPTPVYRWTVDGOEVR 120
Db      139  WVGDNTPQFQSNFESKGDITTYDRGEFHGVDPYDK-FHNTTLDAMDKTTWYLDGESVR 197
QY      121  KTEGGQVSNLTGTG-----LRFNLMS--ESAA---WV-GQFDESKLPLFOFINW 165
Db      198  -----VLSNTSSEGYQSPMYLMMGIWAGGDPNAGTIEMAGETINYNDAFTMYIEK 251
QY      166  VKVYKTYPGQEGGSDFTLMDTDFDGSRMGKCD 202
Db      252  VIVTDYSTGKKYTYGDQGSW-ESIEADGGSIVGRYD 287

```

RESULT 15

```

S34804
exok protein - Rhizobium meliloti megaplasmid 2
C:Species: Rhizobium meliloti
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 26-Aug-1999
C:Accession: S34804; S30072
R:Becker, A.; Kleichmann, A.; Arnold, W.; Puehler, A.
Mol. Gen. Genet. 238, 145-154, 1993
A:Title: Analysis of the Rhizobium meliloti exoX/exoK/exoL fragment: ExoK shows homology
A:Reference number: S34803; MUID:93241147
A:Accession: S34804
A:Molecule type: DNA
A:Residues: 1-268 <BEC>
A:Cross-references: EMBL:Z17219
A:Note: the authors translated the codon GTA for residue 208 as Leu
C:Genetics:
A:Gene: exok
A:Genome: plasmid
C:Superfamily: licheninase

```

```

Query Match      12.5%; Score 180; DB 2; Length 268;
Best Local Similarity 33.8%; Pred. No. 9e-08;
Matches 46; Conservative 26; Mismatches 52; Indels 12; Gaps 6;

```

```

QY      20  YGKFEARMKMAASGIVSSMFLYONGSEIADGRWVEVDIEVLGNPQSFQSN-IITGKA 78
Db      104  YGYEARIKADSGSLNSAFETYIGP--ADKKPHDEIDEEVLGNKTAQVQINQVVSAGK 160
QY      79  GAQKTSKHHAVSPADQAFHTYGLKMTPTPVYRWTVDGOEVRK-TEGGQVSNLTGTGLR 137
Db      161  GNERLAD----VEGGANGENDVAFVWEKNRIRIYVNGELVHEVTDPAKIP--VNAQKIF 214
QY      138  FNLMSSSAA-WVGQF 152
Db      215  FSLWGTFDTLDMWGTF 230

```

Search completed: August 23, 2001, 11:15:07
 Job time: 94 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:15:07 ; Search time 32.87 Seconds

(without alignments)
808.790 Million cell updates/sec

Title: US-09-654-652A-3
Perfect score: 1824
Sequence: 1 MNIKRTAVKSAALAVAAAAA.....AKGAKVNPNGHKRYVNFEPH 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1824	100.0	349	2 A44507	licheninase (EC 3. endo-beta-1,3-1,4
2	268.5	14.7	851	2 H84053	licheninase (EC 3. licheninase (EC 3.
3	265.5	14.6	252	2 A48378	licheninase (EC 3. licheninase (EC 3.
4	248	13.6	242	1 LKBS	licheninase (EC 3. licheninase (EC 3.
5	247.5	13.6	276	2 140453	licheninase (EC 3. licheninase (EC 3.
6	242.5	13.3	239	1 A29091	licheninase (EC 3. xylanase, beta(1,3
7	242.5	13.3	802	2 A36910	licheninase (EC 3. licheninase (EC 3.
8	237	13.0	334	1 S23498	licheninase (EC 3. licheninase (EC 3.
9	233.5	12.8	243	1 S15388	licheninase (EC 3. probable membrane
10	230	12.6	507	2 S64507	licheninase (EC 3. licheninase (EC 3.
11	228	12.5	237	1 S11927	licheninase (EC 3. endo-1,3(4)-beta-g
12	227.5	12.5	238	1 S19012	licheninase (EC 3. hybrid-endo-beta-1
13	224.5	12.3	242	2 J50611	licheninase (EC 3. UTR2 protein - yea
14	190.5	10.4	302	2 G84053	endo-1,3(4)-beta-1
15	189.5	10.4	467	2 S30839	licheninase (EC 3. xylotriose
16	180.5	9.9	268	2 S34804	endo-1,3(4)-beta-1
17	173	9.5	282	2 T02354	endo-1,3(4)-beta-1
18	173	9.5	642	2 B72428	endo-1,3(4)-beta-1
19	172.5	9.5	277	2 B85354	endo-1,3(4)-beta-1
20	170.5	9.3	282	2 A85354	endo-1,3(4)-beta-1
21	164	9.0	286	2 S71225	endo-1,3(4)-beta-1
22	159.5	8.7	287	2 T04236	endo-1,3(4)-beta-1
23	157.5	8.6	269	2 S61555	endo-1,3(4)-beta-1
24	154.5	8.5	277	2 S71222	endo-1,3(4)-beta-1
25	149	8.2	289	2 T06166	endo-1,3(4)-beta-1
26	146.5	8.0	284	2 T52097	endo-1,3(4)-beta-1
27	145.5	8.0	310	2 A86239	endo-1,3(4)-beta-1
28	145	7.9	305	2 G84568	endo-1,3(4)-beta-1
29	140.5	7.7	280	2 T02090	endo-1,3(4)-beta-1

30	139.5	7.6	286	2 T06202	xylotriose
31	138	7.6	422	2 S48564	endo-1,3(4)-beta-1
32	137	7.5	292	2 T06201	endo-1,3(4)-beta-1
33	133.5	7.3	286	2 S48201	endo-1,3(4)-beta-1
34	132	7.2	295	2 T10210	endo-1,3(4)-beta-1
35	129.5	7.1	1324	2 T18265	endo-1,3(4)-beta-1
36	128.5	7.0	277	2 J50156	endo-1,3(4)-beta-1
37	128.5	7.0	283	2 T04514	endo-1,3(4)-beta-1
38	128	7.0	283	2 T07678	endo-1,3(4)-beta-1
39	127.5	7.0	269	2 T05895	endo-1,3(4)-beta-1
40	127	7.0	284	2 T06200	endo-1,3(4)-beta-1
41	125.5	6.9	299	2 D84519	endo-1,3(4)-beta-1
42	123	6.7	296	2 C49539	endo-1,3(4)-beta-1
43	121.5	6.7	292	2 T10211	endo-1,3(4)-beta-1
44	121	6.6	289	2 F71402	endo-1,3(4)-beta-1
45	121	6.6	293	2 T10523	endo-1,3(4)-beta-1

ALIGNMENTS

RESULT 1
A44507
licheninase (EC 3.2.1.73) - Fibrobacter succinogenes
C:Species: Fibrobacter succinogenes
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 08-Oct-1999
C:Accession: A44507
R:Teather, R.M.; Erfle, J.D.
J. Bacteriol. 172, 3837-3841, 1990
A:Title: DNA sequence of a Fibrobacter succinogenes mixed-linkage beta-glucanase (1,3
A:Reference number: A44507; MUID:90299807
A:Accession: A44507
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1349 <TEA>
A:Cross-references: EMBL:M33676; NID:9148575; PIDD:AAA24896.1; PID:9148576
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 100.0% Score 1824; DB 2; Length 349;
Best Local Similarity 100.0%; Pred. No. 7.6e-123;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MNIKRTAVKSAALAVAAAAAALTTNYSADFGSAELTYLEVOYGFEMKMAASGVTS 60
DB 1 MNIKRTAVKSAALAVAAAAAALTTNYSADFGSAELTYLEVOYGFEMKMAASGVTS 60
OY 61 SMFLYONGSEIADGRPWVEVDIEVLGKNPGSFQSNITGKAGAOQTSEKHAHVSPADQA 120
DB 61 SMFLYONGSEIADGRPWVEVDIEVLGKNPGSFQSNITGKAGAOQTSEKHAHVSPADQA 120
OY 121 PFTYCLEMTPNPNVVRWTVGOEVRKTEGGQVSNLTGTGRLFMSSSAAMVGOPEDESKL 180
DB 121 PFTYCLEMTPNPNVVRWTVGOEVRKTEGGQVSNLTGTGRLFMSSSAAMVGOPEDESKL 180
OY 181 PLFQFINWVKYKYPGCGEGSGDFTLDWTFDFDGSRMCKGWTDPGNRVDTLDKNI 240
DB 181 PLFQFINWVKYKYPGCGEGSGDFTLDWTFDFDGSRMCKGWTDPGNRVDTLDKNI 240
OY 241 YSRDGMILALTRKQESFNGQVPRDDEPAPQSSSSAPASSSVPASSSSVPASSSAFV 300
DB 241 YSRDGMILALTRKQESFNGQVPRDDEPAPQSSSSAPASSSVPASSSSVPASSSAFV 300
OY 301 PSSSSATNAIHGMRTTPAVAKEHNLVNAKAKVNPNGHKRYRVNFEPH 349
DB 301 PSSSSATNAIHGMRTTPAVAKEHNLVNAKAKVNPNGHKRYRVNFEPH 349
RESULT 2
H84053
endo-beta-1,3-1,4 glucanase (licheninase) bglIS [imported] - Bacillus halodurans (stra
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 01-Dec-2000

C:Accession: H84053
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A:Reference number: A83650, MUID:20263314
 A:Accession: H84053
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-851 <STO>
 A:Cross-references: GB:AP001518; GB:BA000004; NID:910175792; PIDN:BA06951.1; GSPDB:GN0C
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: bg15

Query Match 14.7% Score 268.5; DB 2; Length 851;
 Best Local Similarity 29.4%; Pred. No. 1.9e-11;
 Matches 83; Conservative 32; Mismatches 112; Indels 55; Gaps 12;

21 LTTNVAKDFGAELELYTEVYQKFEARKMAAGSTVSMFLYONGSEIADGRPWVEV 80
 93 LENEQAPRYAGELRTNDFYQYGLFEVSMKPAVEGVTSFFTY-TCGEMWDGDPNDEI 151
 81 DIEVLGNPGSEFGSNITTKAGAKTSEKHNHVPADDAFHTYGLEWTPVYKWTYDQ 140
 152 DIEFLGDTTRIQFNFTNGVG--NFTYDLGFDASESFNTYAFEMREDSITVYNGE 208
 141 EVRTBEGOVSNLTGT-OGLEFNLMSSSA-AWGOFDSESKLPLFOFINVYKYKTP-- 196
 209 AVHTA---TENIPOTPKIMNLMVPGVDEMGVDEGDTNPHADYEWV---RYTPLE 261
 197 --GQEGSGPFTLDMT-----DNFTDFDSRMC-KGDMTFDG 230
 262 ALDEESGDDEEPEVEEPADEDEVSRPGSAIYEFNFNEDINSIANGWI-NG 320
 231 NRDLT--DKNIYSRDCMLIALTRKQSEFNGVPRDDEPA 270
 321 QMFNATWYDSQIFPSNGIMRPAIDK-----DDEEA 351

RESULT 3
 A48378
 A:Accession: EC 3.2.1.73 - *Bacillus brevis*
 N:Alternate names: lichenase
 C:Species: *Bacillus brevis*
 C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 07-May-1999
 C:Accession: A48378
 R:Low, M.E.; Reid, S.J.; Watson, T.G.
 Appl. Microbiol. Biotechnol. 38, 507-513, 1993
 A:Title: Characterization, cloning and sequencing of a thermostable endo-(1,3-1,4) beta-
 A:Reference number: A48378; MUID:93159752
 A:Accession: A48378
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-252 <DOU>
 A:Note: sequence extracted from NCBI backbone (NCBI:124672, NCBI:124673)
 C:Superfamily: lichenase
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 14.6% Score 265.5; DB 2; Length 252;
 Best Local Similarity 35.8%; Pred. No. 6.7e-12;
 Matches 62; Conservative 24; Mismatches 74; Indels 13; Gaps 6;
 26 SAUDFSAELYTEVYQKFEARKMAAGSTVSMFLYONGSEIADGRPWVEV 85
 90 SANRYKAGELRTNDFYQYGLFEVSMKPAVEGVTSFFTY-TCGEMWDGDPNDEI 148
 86 GKNPGFSNITTKAGAKTSEKHNHVPADDAFHTYGLEWTPVYKWTYDQ 145
 149 GKPTTRQFNFTNGVG--NFTYDLGFDASESFNTYAFEMREDSITVYNGEAVHTA 205
 146 EGGOVSNLTGT-OGLEFNLMSSSA-AWGOFDSESKLPLFOFINVYKYKTP 196

Db 206 ----TENIPOTPKIMNLMVPGVDEMGVDEGDTNPHADYEWV---RTP 251

RESULT 4
 LXB5
 A:Accession: EC 3.2.1.73 precursor - *Bacillus subtilis*
 N:Alternate names: 1,3-1,4-beta-D-glucan 4-gluconohydrolase; beta-glucanase; lichenase
 N:Contents: lichenase E-1; lichenase E-2
 C:Species: *Bacillus subtilis*
 C:Date: 30-Jun-1988 #sequence_revision 13-Mar-1998 #text_change 16-Jun-2000
 A:Accession: I40370; B69594; A22914; A90026; A90027; JU0110
 R:Wolf, M.; Geczi, A.; Simon, O.; Borris, R.
 Microbiology 141, 281-290, 1995
 A:Title: Genes encoding xylan and beta-glucan hydrolysing enzymes in *Bacillus subtilis*
 A:Reference number: I40370; MUID:95219081
 A:Accession: I40370
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-242 <RES>
 A:Cross-references: EMBL:246862; NID:9599673; PIDN:CAA86922.1; PID:9599674
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.: Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.: Ehrlich, S.D.; Emmerison, P.T.; Entian, K.D.; Erington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritzt, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gal
 lech, J.; Harwood, C.R.; Henault, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kuman, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portele
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowski, A.; Se
 kuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Witters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunaka, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, K.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
 A:Reference number: A69580; MUID:98044033
 A:Accession: B69594
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-242 <KUN>
 A:Cross-references: GB:299124; GB:AL009126; NID:92636442; PIDN:CAB15943.1; PID:926364
 R:Murphy, N.; McConnell, D.J.; Cantwell, B.A.
 Nucleic Acids Res. 12, 5355-5367, 1984
 A:Title: The DNA sequence of the gene and genetic control sites for the excreted B. s
 A:Reference number: A93526; MUID:84272222
 A:Accession: A22914
 A:Molecule type: DNA
 A:Residues: 1-203, 'L', 205-242 <MUR>
 A:Cross-references: EMBL:X00754; NID:939818; PIDN:CAA25328.1; PID:9685236
 A:Experimental source: strain C120
 A:Note: the authors translated the codon CAA for residue 29 as Lys and CCA for residu
 R:Tezuka, H.; Yuuki, T.; Yabuuchi, S.
 Agric. Biol. Chem. 53, 2335-2339, 1989
 A:Title: Construction of a beta-glucanase hyperproducing *Bacillus subtilis* using the
 A:Reference number: A90026
 A:Accession: A90026
 A:Molecule type: DNA
 A:Residues: 1-23, 'S', 25-82, 'S', 84-242 <TEZ>
 A:Cross-references: DBJ:D00518; NID:9216243; PIDN:BA00405.1; PID:9216244
 R:Yuuki, T.; Tezuka, H.; Yabuuchi, S.
 Agric. Biol. Chem. 53, 2341-2346, 1989
 A:Title: Purification and some properties of two enzymes from a beta-glucanase hyperp
 A:Reference number: A90027
 A:Contents: annotation
 A:Note: source was hyperproducing strain HL-25 with gene from strain Y-25
 C:Genetics:
 A:Gene: bg15
 C:Function:
 A:Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in lichenin an

R:Schlaming, S.; Schwarz, W.H.; Staudenbauer, W.L.
Eur. J. Biochem. 204, 13-19, 1992
A:Title: Structure of the Clostridium thermocellum gene l1cB and the encoded beta-1,3-1,
l cellulases.
A:Reference number: S23498; MUID:92155194

A:Accession: S23498
A:Molecule type: DNA
A:Residues: 1-334 <SCH>
A:Cross-references: EMBL:X63355; NID:g40697; PIDN:CAA44959.1; PID:g40698
C:Genetics:

A:Gene: l1cB
C:Superfamily: licheninase l1cB; Clostridium cellulase repeat homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-334/Product: licheninase #status predicted <MNT>
F:273-296/Domain: Clostridium cellulase repeat homology <CCRI>
F:308-331/Domain: Clostridium cellulase repeat homology <CCR2>

Query Match 13.0%; Score 237; DB 1; Length 334;
Best Local Similarity 30.9%; Pred. No. 1e-09;

Matches 80; Conservative 27; Mismatches 94; Indels 58; Gaps 13;

QY 28 KDFSGAELTYLEVO-----YGFKFAKMAAAGTVSSMFLYQNGSEIADGRPWVEVD 81
DB 82 REYGSYPYKSGEYRTKSFEGYGEYVEMAKAKNVGIVSFFTYGPS---DNMPWDEID 138
QY 82 IEVIGKKNPGSFOSNITIGKGAOKTSEKHNHVAAPADAFHTYGLFEMTPNVYRTVTCOE 141
DB 139 IEFLGKDTTKVQFWMYKNGVGG---NEYLHNLGFDASODFTTYGFEMPPDIDIFYVGGKK 195
QY 142 VRKTEGQVSNLTGTGQ-LRFNLWSSSA-AWQGFDESKLPLOFQIMVAVY----- 192
DB 196 VYR-----GTRNIPYTPCKIMNMLPVGIGVDMLGRYD-GRTPLOAEYEVYVYRNGVPQD 250
QY 193 -----KYP-----GQGE-GGSDFT-----LDWTDNFDTEFGSSRMKG 224
DB 251 NPPTPTIAPTPTNPMLPLKGVNCGDHVNSDSYLFKRYLLRVIDRFPVGDQS---VA 307
QY 225 DWTFDGNRVLDLTKNIYSR 243
DB 308 DVNRDG-RIDSTDLMLKR 325

RESULT 9

l1cB: licheninase (EC 3.2.1.73) - Bacillus licheniformis
N:Alternate names: endo-beta-1,3-1,4-D-glucanase; lichenase
C:Species: Bacillus licheniformis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S15388
R:Roberts, J.; Perez-Pons, J.A.; Querol, E.
Eur. J. Biochem. 197, 337-343, 1991
A:Title: Molecular cloning, expression and nucleotide sequence of the endo-beta-1,3-1,4-
A:Reference number: S15388; MUID:91224124
A:Accession: S15388
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-243 <LIO>
A:Cross-references: EMBL:X57279; NID:g39558; PIDN:CAA0547.1; PID:g39559
C:Superfamily: licheninase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 12.8%; Score 233.5; DB 1; Length 243;
Best Local Similarity 33.7%; Pred. No. 1.2e-09;

Matches 58; Conservative 26; Mismatches 77; Indels 11; Gaps 6;

QY 23 TNVSAKDFSGAELTYLEVOYKGFARMAAAGTVSSMFLYQNGSEIADGRPWVEVD 82
DB 81 TSPSYNKFDCGKNSVPTTYGILEVNMKPAKKNVGVSSFFTYGPT---DGFPMDEID 137
QY 83 EVLGNKPGSFOSNITIGKGAOKTSEKHNHVAAPADAFHTYGLFEMTPNVYRTVTCOE 142

DB 138 EFLGKDTTKVQFNYYTNGVG---NHEKIVNLGFDAAANSYHTYADMQPNSIKRWVDCQ-L 193
QY 143 KRTGGOVSNLTGTGQLRFNLWSSSA-AWQGFDESKLPLOFQIMVAVYK 193
DB 194 KHTATQIPIQTPGK--IMMNLNMGAGVDEWLGSYN-GVTPLSRLHMYRYTK 242

RESULT 10

S64507
probable membrane protein YGR189c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein G7553
C:Species: Saccharomyces cerevisiae
C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 29-Oct-1999
C:Accession: S64507
R:Arroyo, J.; Garcia-Gonzalez, M.; Garcia-Saez, M.I.; Sanchez-Perez, M.; Nombela, C.
Submitted to the Protein Sequence Database, May 1996
A:Reference number: S64499
A:Accession: S64507
A:Molecule type: DNA
A:Residues: 1-507 <ARR>
A:Cross-references: EMBL:Z72974; NID:g1323335; PID:e243566; PID:g1323336; GSPDB:GN000
A:Experimental source: strain S288C
C:Genetics:
A:Gene: MIPS:YGR189c
A:Map position: 7R
C:Keywords: transmembrane protein
F:6-22/Domain: transmembrane #status predicted <TM>

Query Match 12.6%; Score 230; DB 2; Length 507;
Best Local Similarity 26.9%; Pred. No. 5.6e-09;

Matches 87; Conservative 57; Mismatches 134; Indels 46; Gaps 13;

QY 6 TAVKALAAVAAAALTTNWSAKDFSGAELTYLEVOYKGFARMAAAGTVSSMFLY 65
DB 70 TDLKHAIEIKYSGSLGSLMTL-AKRYDNPISLKSNEYIMYKGLVILKAACTGIVSSFYLO 128
QY 66 QNGSEIADGRPWVEDEVIGKKNPGSFOSNITIGKGAOKTSEKHNHVAAPADAFHTY 125
DB 129 SDLDL-----EIDLEWVGDTQFOSNFSKGTDTTYDRGEFFGVPTDR-FHNYT 179
QY 126 LEWTPNVYRTVDCQEVKRTGQVSNLTGTGQ-----LRFNLWSS--ESAA---W 171
DB 180 LDWAMDKTWYLDGSEVR-----VLSNTSSEGVPOSPMLVMGIMVAGPDMAAGTIEN 233
QY 172 V-GQFDESKLPLOFQIMVAVYKTPGQEGSGDFTLDWTDNFDTEFGSRMGKDWTFDG 230
DB 234 AGGETYNDAFTMYIEKIVTVDTSTGKRYTYGQSGSW-ESIEADGSIYGRYDQAE- 291
QY 231 NRVDLTGKNIYSRDMILALTRKG--QESFNGOVPRDDEPAPOSSSSAPASS--SSVPA 286
DB 292 -----DFAVLANGSISSSSTSSSTYSASSASTVSSSVSSASTVSSSVSTVS 345
QY 287 SSSSVPASSSSAFLVPPSSSATNA 310
DB 346 SSSSVSSSSSTS---PSSSTATSS 366

RESULT 11

S11927
l1cB: licheninase (EC 3.2.1.73) precursor [validated] - Bacillus macerans
N:Alternate names: endo-beta-1,3-1,4-glucanase; l1cB: lichenase
C:Species: Bacillus macerans
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 15-Sep-2000
C:Accession: S11927
R:Boeris, R.; Bueltner, K.; Maentzsaalae, P.
Mol. Gen. Genet. 222, 278-283, 1990
A:Title: Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans: homologie
A:Reference number: S11927; MUID:91109712
A:Accession: S11927
A:Status: preliminary
A:Molecule type: DNA

Query Match	12.5%	Score 228	DB 1	Length 237
Best Local Similarity	34.3%	Pred. No. 2.9e-09		
Matches 60; Conservative	25;	Mismatches 74;	Indels 16;	Gaps 8;

Db 188 KHT--ATANIPSTPGKIMMNLWNGTGVDLDLGSYNGAN-PLYAEYDWV--KYT 235

C:\Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 18-Jun-1999

A;Reference number: S19011; MUID:92041687

```

F;129/Active site: Glu #status predicted

```

Matches	61;	Conservative	26;	Mismatches	74;	Indels	17;	Gaps	9;
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

Db 74 SLTSPANKK-FDCGEYRSTNNYGYGLYEVSMPAKNTGIVSSFFTYTGPSh--GTQWDE 129

Db 187 -VLKHT--ATTNIPSTPGKIMMNLWNGTGVDSWLGSYNGAN-PLYAEYDWV---KYT 236

RESULT 13

A;Reference number: JS0611; MUID:92095946

A;Residues: 1-242 <2

R; Zverlov, V.V.; Velikodvorskaya, G.A

Biotechnol. Lett. 12, 811-816, 1990

A;Reference number: S18726

A; Accession: S18726

A: molecule type: DNA

A;Residues: 1-242 <ZVE2>

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1991

C; Genetics:

h/scie: fungi
C;Superfamily: Licheninase

C;Keywords: glycosidase; h

F:/28-242/Product: endo-1,3(4)-beta-glucanase #status predicted <END>

T O U S : 000001/000002, 000003

Matches	62;	Conservative	25;	Mismatches	71;	Indels	27;	Gaps	8;
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

QY 14 VAAAAAALTNNVS-----AKDFSGAELTYLLEEVQ-----YCKFEARMKMAAASGTV 59

[illegible][illegible]

00 35ME L1 QNGSEIADONFWEADIE LGNNFOSF QSN L1 GNNFONJ GNNNNV G FNNY 11

Db 118 SSEFTYTGP

120 AFHTYGLEWTPNYVRWTVDGQEVKRTKEGGOVSNLTGTQG-LRFNLWSSESA-AWVGQFDE 177

18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
8

DD I / 2 DEMILITARY DIVISION OF THE ARMY

178 SKLPL 182

Db 227 GRTPL 231

DD FORM 427 6/15/64

RESULT 14

RESUL 14
G84053

C:Species: *Bacillus halodurans*
 glycid-endo-beta-1,3-1,4 glucanase bhz01 (imported) *Bacillus halodurans* (strain C)

C:\Date: 01-Dec-2000 #sequence

C;Accession: G84053
R;Takami, H.: Nakasone, K.: Takaki, Y.: Maeno, G.: Sasaki, R.: Masui, N.: F

Nucleic Acids Res

A; Title: complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a

A:Reference number: A83650; MUID:20263314
A:Accession: GB4053
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-302 <SMD>
A:Cross-references: GB:AP001518; GB:BA000004; NID:g10175792; PIDN:BAB06950.1; GSPDB:GNOC
A:Experimental source: strain C-125
C:Genetics:
.Gene: BH3231

Query Match	10.4%	Score 190.5	DB 2	Length 302
Best Local Similarity	27.8%	Pred. No. 1.9e-06		
Matches	62	Conservative	34	Mismatches 98
				Indels 29
				Gaps 9

QY 29 DFGSELVLTLEVOYGRFPEARMKAAASGYSSMELQNSSELADGRPNVLEIVLGKN 88
 Db 16 NYECAESTHNFYTGLEIVSMKPSNVSGLVSIFFYTGPS--YNGAPMEIDIEFLGND 73
 QY 89 PGSFOSNIITGKAGAOKTSEKHNAVSPAADAFHTYTGLEMTPNVVRATVYDGOEVRKTEGG 148
 Db 74 TTQVQPNVYTINGRG--NELYLDLGRDANSPFTYAFMDQENHTINNYVAGSELVATATEN 130
 QY 149 QVSNLTGTQGLREPLWSE-----SAAWVGQFDESKLPQLFQITNWKVYKYP--GQG 199
 Db 131 IPSN--PSKMMNMNIWNTYCIDEWAGAYWCGONANAS-----YEWV---RYTPNNERSQT 178
 QY 200 EGGSDFTLDWTDNFDTDGSM--GKDWTFDGRVULDTOKNI 240
 Db 179 PIASDFOLHACEYSIDARGVYSWQGVGSF--YPGSMIFDVNVL 220

RESULT 15

530839

UTR2 protein - yeast (*Saccharomyces cerevisiae*)
N/Alternate names: protein YEL040w

C;Species: *Saccharomyces cerevisiae*

C; Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 06-Feb-1998

```
C:\Accession: $30839: $50504: $38545
C:\Date: 28-May-1993 #sequence_rev15
```

C:\AccessIon: S30839; S50504; S38545
B: Mulligan, J T : Dietrich, F S : Hennessey, K M : Schi, D : Koen, C : Woj, Y : Marshall

R; Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor, submitted to the EMU Data Library February 1993

submitted to the EMBL Data

A; Reference number:

A; Accession: S30839

A; molecule type: DNA

A;Residues: 1-4

A:Cross-references: GB:U18779: EMBL:L110830: NID:α
A;Residues: 1-40/ <MOL>

A;Cross-references: GB:U18779; EMBL:L10830; NID:g603625; PID:g603639
R:Dierrich F S

R; Dietrich, F.S.

submitted to the EM

A;Description: The s

A; Reference number: S50

Accession: S50504

A; molecule type: DNA
A; accession: 520204

```
A:ResIdues: 1-467 <DTE>
```

A;Residues: 1-467 <DIE>
A:Cross-references: EMBL:U18770. NID:5603635. RID:5603630. WISC-VER 040.

A;C/Cross-references: EMBL:U18779; NID:g603333

R; Melnick, L.; Sher

J. Mol. Biol. 233, 3

A;Title: The gene clusters ARC and COR on chromosomes 5 and 1

A; Reference number: S38543; MUID:94016558

A; Accession

A:MOLECULE TYPE

A: Molecule type: DNA
A: Best fit: 121-128 /V/ 121-200 /B/ 203-252 /C/

A;Residues: 121-12

A:Cross-references: EMBL:SS6130; NID:g430829; PID:g430830
C:Gene: SGD:UTR2
A:Cross-references: SGD:S0000766; MIPS:YEL040w
A:Map position: 5L

Search completed: August 23, 2001, 11:15:08
Job time: 95 sec

[illegible]

Matches 349: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKKTAVKALVAAAAAALTTNSKDFSGAELYLEVOYGFKFAKMAASTVS 60
 1 MNKKTAVKALVAAAAAALTTNSKDFSGAELYLEVOYGFKFAKMAASTVS 60
 Db 1 MNKKTAVKALVAAAAAALTTNSKDFSGAELYLEVOYGFKFAKMAASTVS 60

QY 61 SMFLYONGSEIADGRPWVEVDIEVLGNKPSFOSNITTKAGAOKTSEKHHAVSPADQA 120
 61 SMFLYONGSEIADGRPWVEVDIEVLGNKPSFOSNITTKAGAOKTSEKHHAVSPADQA 120
 Db 61 SMFLYONGSEIADGRPWVEVDIEVLGNKPSFOSNITTKAGAOKTSEKHHAVSPADQA 120

QY 121 FHNYGLEMTNRYRWYVDGGEVKTGEGOVSNLTGDLGFNLMSSSAAMVGOFPDESKL 180
 121 FHNYGLEMTNRYRWYVDGGEVKTGEGOVSNLTGDLGFNLMSSSAAMVGOFPDESKL 180
 Db 121 FHNYGLEMTNRYRWYVDGGEVKTGEGOVSNLTGDLGFNLMSSSAAMVGOFPDESKL 180

QY 181 PLFOFIMVWVYVYTPQGGSGSDFLTMDTNDPTFDGSRWKGDMTFDGNRYDLTKNI 240
 181 PLFOFIMVWVYVYTPQGGSGSDFLTMDTNDPTFDGSRWKGDMTFDGNRYDLTKNI 240
 Db 181 PLFOFIMVWVYVYTPQGGSGSDFLTMDTNDPTFDGSRWKGDMTFDGNRYDLTKNI 240

QY 241 YSRDGMILIALTRKGOSEFNGOVRPDEPAPQSSSSAPASSSVPAASSSSAFV 300
 241 YSRDGMILIALTRKGOSEFNGOVRPDEPAPQSSSSAPASSSVPAASSSSAFV 300
 Db 241 YSRDGMILIALTRKGOSEFNGOVRPDEPAPQSSSSAPASSSVPAASSSSAFV 300

QY 301 PSSSSATNATIHGMRTTPAVAKERHLYNKAQKVPNGKRRVNEFH 349
 301 PSSSSATNATIHGMRTTPAVAKERHLYNKAQKVPNGKRRVNEFH 349
 Db 301 PSSSSATNATIHGMRTTPAVAKERHLYNKAQKVPNGKRRVNEFH 349

RESULT 2
 GUB_BACBR STANDARD: PRT: 259 AA.

AC P37073;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DE 01-JUN-1994 (Rel. 29, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 GN (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
 OS BGLB.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Brevibacillus.
 OX NCBI_TaxID=1393;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93159752; PubMed-7763386;
 RA Louw M.E., Reid S.J., Watson T.G.;
 RT "Characterization, cloning and sequencing of a thermostable
 RT endo-(1,3-1,4) beta-glucanase-encoding gene from an alkalophilic
 RT Bacillus brevis";
 RL Appl. Microbiol. Biotechnol. 38:507-513(1993).
 CC -1- FUNCTION: HYDROLYSES B-GLUCANS CONTAINING MIXED BETA-1,3 AND
 CC BETA-1,4 LINKAGES. THIS ENZYME IS THERMOSTABLE; ITS OPTIMAL
 CC TEMPERATURE IS NEAR 65-70 DEGREES CELSIUS.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
 CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
 CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
 CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC EMBL: M84339; AAA22265.1; -
 DR HSSP: P23904; IAJK.
 DR InterPro: IPR000757; -
 DR Pfam: PF00722; Glyco_hydro.16; 1.
 DR PRINTS: PR00737; GLHYDRLASE16.
 DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
 KW Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 31 POTENTIAL.
 FT CHAIN 32 259 BETA-GLUCANASE.
 FT ACT SITE 142 142 NUCLEOPHILE (BY SIMILARITY).
 FT ACT SITE 146 146 PROTON DONOR (BY SIMILARITY).
 SQ SEQUENCE 259 AA; 23960 MM; A630C9F281FF5D13 CRC64;

Query Match 14.6%; Score 266.5; DB 1; Length 259;
 Best Local Similarity 36.0%; Pred. No. 3,9e-12;
 Matches 63; Conservative 24; Mismatches 75; Indels 13; Gaps 6;

QY 26 SAKDFSGAELYLEVOYGFKFAKMAASTVS SMFLYONGSEIADGRPWVEVDIEVL 85
 26 SAKDFSGAELYLEVOYGFKFAKMAASTVS SMFLYONGSEIADGRPWVEVDIEVL 85
 Db 26 SAKDFSGAELYLEVOYGFKFAKMAASTVS SMFLYONGSEIADGRPWVEVDIEVL 85

QY 90 SARNYKAGELRTNDFEYHYGLFEVSMKPAKVEGTVSFFTY-TGEMDMGDPMDIEFL 148
 90 SARNYKAGELRTNDFEYHYGLFEVSMKPAKVEGTVSFFTY-TGEMDMGDPMDIEFL 148
 Db 90 SARNYKAGELRTNDFEYHYGLFEVSMKPAKVEGTVSFFTY-TGEMDMGDPMDIEFL 148

QY 86 GRNPGSQSNITTKAGAOKTSEKHHAVSPADQA FHTGTGLETTPYVRYVDGGEVKT 145
 86 GRNPGSQSNITTKAGAOKTSEKHHAVSPADQA FHTGTGLETTPYVRYVDGGEVKT 145
 Db 86 GRNPGSQSNITTKAGAOKTSEKHHAVSPADQA FHTGTGLETTPYVRYVDGGEVKT 145

QY 149 GKDTRIQFYFTNGVGG--NEFYDGLGFDASESFNTAFEFWRREDSITWYVNGEAVHTA 205
 149 GKDTRIQFYFTNGVGG--NEFYDGLGFDASESFNTAFEFWRREDSITWYVNGEAVHTA 205
 Db 149 GKDTRIQFYFTNGVGG--NEFYDGLGFDASESFNTAFEFWRREDSITWYVNGEAVHTA 205

QY 146 EGGQVSNLTGT-GLRPNLMSSSA-AWVGQFPDESKLPLFOFIMVWVYVYTPQ 198
 146 EGGQVSNLTGT-GLRPNLMSSSA-AWVGQFPDESKLPLFOFIMVWVYVYTPQ 198
 Db 146 EGGQVSNLTGT-GLRPNLMSSSA-AWVGQFPDESKLPLFOFIMVWVYVYTPQ 198

QY 206 -----TENIPOTPOKIMNLMFGVGVGDMTCVFDGDMTPYVSYDWT---RTTPLQ 253
 206 -----TENIPOTPOKIMNLMFGVGVGDMTCVFDGDMTPYVSYDWT---RTTPLQ 253
 Db 206 -----TENIPOTPOKIMNLMFGVGVGDMTCVFDGDMTPYVSYDWT---RTTPLQ 253

RESULT 3
 GUB_BACSU STANDARD: PRT: 242 AA.

AC P04957;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
 GN BGLS OR BGL OR LICS OR N15B.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-C120;
 RX MEDLINE-84272222; PubMed-6087283;
 RA Murphy N., McConnell D.J., Cantwell B.A.;
 RT "The DNA sequence of the gene and genetic control sites for the
 RT excreted B. subtilis enzyme beta-glucanase";
 RL Nucleic Acids Res. 12:5355-5367(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN-HL-25;
 RC Tezuka H., Yuuki T., Yabuuchi S.;
 RA "Construction of a beta-glucanase hyperproducing Bacillus subtilis
 RT using the cloned beta-glucanase gene and a multi-copy plasmid";
 RL Agric. Biol. Chem. 53:2335-2339(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN-168 / BGSC1A1;
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
 RL Submitted (FEB-1995) to the EMBL/Genbank/DBD databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RX STRAIN-168;
 RC MEDLINE-95219081; PubMed-7704256;
 RA Wolf M., Geczi A., Simon O., Boriss R.;
 RT "Genes encoding xylan and beta-glucan hydrolysing enzymes in Bacillus
 RT subtilis: characterization, mapping and construction of strains
 RT deficient in lichenase, cellulase and xylanase";
 RL Microbiology 141:281-290(1995).
 RN [5]
 RP SEQUENCE OF 1-6 FROM N.A.
 RX STRAIN-BR151;
 RX MEDLINE-96178961; PubMed-8606172;
 RA Schmetz K., Stuelke J., Getz S., Krueger S., Krieger M., Hecker M.,
 RA Rak B.;

```

RT "Lact, a Bacillus subtilis transcriptional antiterminal protein of
RT the BglG family."
RT J. Bacteriol. 178:1971-1979(1996).
RN [6]
RP SEQUENCE OF 29-63.
RA Yuuki T., Tezuka H., Yabuuchi S.;
RT "Purification and some properties of two enzymes from a beta-glucanase
RT hyperproducing strain, Bacillus subtilis HL-25."
RL Agric. Biol. Chem. 53:2341-2346(1989).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, X00754; CAA25328.1; -
DR EMBL, D00518; BAA00405.1; -
DR EMBL, D83026; BAA11697.1; -
DR EMBL, Z46862; CAA86922.1; -
DR EMBL, Z28340; CAA82195.1; -
DR EMBL, Z59124; CAA15943.1; -
DR PIR, A22914; LXBS.
DR PIR, J00110; J00110.
DR HSSP, P27051; 1GBG.
DR Subtilist; BG10476; bglS.
DR InterPro; IPR000757; -.
DR Pfam; PF00722; Glyco_hydro_16; 1.
DR PRINTS; PR00737; GLHYDRLASE16.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
DR KW Hydrolyase; Glycosidase; Signal.
FT SIGNAL 1 28
FT CHAIN 29 242
FT MOD_RES 29 29
FT ACT_SITE 133 133
FT ACT_SITE 137 137
FT DISULFID 60 89
FT VARIANT 24 24
FT VARIANT 83 83
FT VARIANT 204 204
SQ SEQUENCE 242 AA; 27268 MW; 45958DEA70F22B29 CRC64;

Query Match 13.6%; Score 248; DB 1; Length 242;
Best Local Similarity 36.0%; Pred. No. 7.3e-11;
Matches 63; Conservative 25; Mismatches 75; Indels 12; Gaps 7;

OY 20 ALTVNSAKDFSGAELTYLEEVYGRKFEARMKMAASGVSMFLYQNSGLADGRPWV 79
DB 78 ALTPAYNVN-FPCGGERSVYQYGYGLYERMRKPAKNTGVSFFYTGPT--DGPWDE 133
OY 80 VDIEVLGNKPGSFQSNILITGKGAQKTSKHHAVSPADQAQHTYGLETPYVWTVGVG 139
DB 134 IDIEFLGDDTTRVQFNYYTINGAG---NHEKIVDLGFDANAHHTAFAFDQPSIKRYVVG 190
OY 140 QEVKRTGEGVSNLTGTQGLRPNLWSSESA-AWVGQFDESKLPDFQFINWVYK 193
DB 191 Q-LKHTATNQLPTTPGK--IMMNLWNGTGVDMLGSYNGVN-PIVAHYDMVRYTK 241

RESULT 4
GUB_BACAM STANDARD; PRT; 239 AA.
AC P07980;

```

```

DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
GN BGLA.
OS Bacillus amyloliquefaciens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BE 20/78;
RX Hofmeister J., Kurtz A., Borris R., Knowles J.;
RT "The beta-glucanase gene from Bacillus amyloliquefaciens shows
RT extensive homology with that of Bacillus subtilis."
RL Gene 49:177-187(1986).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, M15674; AAA87323.1; -
DR PIR, A29091; A29091.
DR HSSP, P27051; 1GBG.
DR InterPro; IPR000757; -.
DR Pfam; PF00722; Glyco_hydro_16; 1.
DR PRINTS; PR00737; GLHYDRLASE16.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
DR KW Hydrolyase; Glycosidase; Signal.
FT SIGNAL 1 25
FT CHAIN 26 239
FT ACT_SITE 134 134
FT DISULFID 57 86
SQ SEQUENCE 239 AA; 26928 MW; A76A64268A7AAA0B CRC64;

Query Match 13.3%; Score 242.5; DB 1; Length 239;
Best Local Similarity 34.3%; Pred. No. 1.8e-10;
Matches 59; Conservative 27; Mismatches 75; Indels 11; Gaps 6;

OY 23 TVNSAKDFSGAELTYLEEVYGRKFEARMKMAASGVSMFLYQNSGLADGRPWV 82
DB 77 TSPSYNKKFPCGGERSVYQYGYGLYERMRKPAKNTGVSFFYTGPT--GTPDEIDI 133
OY 83 EVLGNKPGSFQSNILITGKGAQKTSKHHAVSPADQAQHTYGLETPYVWTVGVG 142
DB 134 EFLGDDTTRVQFNYYTINGAG---NHEKIVDLGFDANAHHTAFAFDQPSIKRYVVG 189
OY 143 RKEGQGVSNLTGTQGLRPNLWSSESA-AWVGQFDESKLPDFQFINWVYK 193
DB 190 KHTATTLQIPAAVGK--IMMNLWNGTGVDMLGSYNGVN-PIVAHYDMVRYK 238

RESULT 5
XEND_RUMFL
ID XEND_RUMFL STANDARD; PRT; 802 AA.
AC Q53317;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE XYLANASE/BETA-GLUCANASE PRECURSOR (INCLUDES: ENDO-1,4-BETA-XYLANASE
DE (EC 3.2.1.8) (XYLANASE); (ENDO-BETA-1,3-1,4 GLUCANASE) (EC 3.2.1.73)

```



```

DR InterPro: IPR000757; -
DR InterPro: IPR002048; -
DR InterPro: IPR002105; -
DR Pfam: PF00404; Dockerin_1; 2.
DR Pfam: PF00722; Glyco_hydro_16; 1.
DR PRINTS: PR00737; GLYHYDRLASE16
DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_2.
DR PROSITE: PS00448; CLOS_CELLOLOSOME_RPT; 2.
KW Hydrolase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 27
FT CHAIN 28 334 BETA-GLUCANASE.
FT ACT_SITE 136 136 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 140 140 PROTON DONOR (BY SIMILARITY).
FT DOMAIN 252 269 PRO/THR-RICH (LINKER).
FT DOMAIN 273 331 2 X 24 AA APPROXIMATE REPEATS.
FT REPEAT 273 296 1.
FT REPEAT 308 331 2.
FT CONFLICT 304 334 QSVADVNRDGRIDSTDLMLKRLIRAIPL ->
FT PODCGRHDRVDSGSK (IN REF. 2 AND 3).
SQ SEQUENCE 334 AA; 37897 MW; 0837564E9726F281 CRC64;

Query Match 13.08; Score 237; DB 1; Length 334;
Best Local Similarity 30.9%; Pred. No. 6,4e-10;
Matches 80; Conservative 27; Mismatches 94; Indels 58; Gaps 13;

QY 28 KDFSGAEIYLTLEVOY-----YGFKEARMKMAASGVSSMFLYONGSEIADGRPWVEVD 81
DB 82 KEYGSGYTKGGEYRTKGFEGFYGVRRKAKNGIVISFFTYTGPS--DNNWMDID 138
QY 82 IEVLGKNPDSFOSNITITKAGAKTSEKHAHVSPADQAFHTYGLDEWPNVYRWTVDSQE 141
DB 139 IELFGKDTTKYQOFMYNKGVG--NEYLHNLGPDASQDFHTYGEKMPDVIDEYVDSKK 195
QY 142 VRKEGGGVSNLTGTQG-LRNLMSSESA-AWVGQFDESKLPLEFQIIMWVKY----- 192
DB 196 VYR---GTRNIPVYPRGIMNMLMPCIGVDEMLGRYD-GRTPLQAEYEVYVYYPNGVPOD 250
QY 193 -----KTRP-----GQGE-GGSDEP-----LMDTMDPDPEDSGRMCKG 224
DB 251 NPPTPTAPSTPTNPNIPLKGDVNGDSGVHNSDSYLFKRYLLRYIDRFPVGDQS---VA 307
QY 225 DWTEFGNREVDLTDKNYSR 243
DB 308 DVNRDG-RIDSTDLMLKR 325

RESULT 7
GUB_BACLI STANDARD; PRT: 243 AA.
ID GUB_BACLI
AC P27051;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHEVASE).
GN BGI.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_Taxid=1402;
RN [1]
RP MEDLINE=91224124; PubMed=2026156;
RA Lloberas J., Perez-Pons J.A., Querol E.;
RT "Molecular cloning, expression and nucleotide sequence of the
RT endo-beta-1,3-1,4-D-glucanase gene from Bacillus licheniformis.
RT Predictive structural analyses of the encoded polypeptide.";
RL Eur. J. Biochem. 197;337-343(1991).
RN [2]
RP REVISIONS.
RA Querol E.;

```

```

RL Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RP MUTAGENESIS.
RX MEDLINE=92362869; PubMed=1354172;
RA Planas A., Juncosa M., Lloberas J., Querol E.;
RT "Essential catalytic role of Glu134 in endo-beta-1,3-1,4-D-glucan 4-
RT glucanohydrolase from B. licheniformis as determined by site-directed
RT mutagenesis.";
RL FEBS Lett. 308:141-145(1992).
RN [4]
RP MUTAGENESIS.
RX MEDLINE=94237863; PubMed=8182059;
RA Juncosa M., Pons J., Dot T., Querol E., Planas A.;
RT "Identification of active site carboxylic residues in Bacillus
RT licheniformis 1,3-1,4-beta-D-glucan 4-glucanohydrolase by
RT site-directed mutagenesis.";
RL J. Biol. Chem. 269:14530-14535(1994).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=96063718; PubMed=7589539;
RA Hahn M., Pons J., Planas A., Querol E., Heinemann U.;
RT "Crystal structure of Bacillus licheniformis 1,3-1,4-beta-D-glucan 4-
RT glucanohydrolase at 1.8-A resolution.";
RL FEBS Lett. 374:221-224(1995).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC -1- IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL: X57279; CAA04547.1; -
DR PIR: S15388; S15388
DR PDB: 1GBG; 07-DEC-95.
DR InterPro: IPR000757; -
DR Pfam: PF00722; Glyco_hydro_16; 1.
DR PRINTS: PR00737; GLYHYDRLASE16.
DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
KW Hydrolase; Glycosidase; Signal; 3D-structure.
FT SIGNAL 1 27
FT CHAIN 28 243 BETA-GLUCANASE.
FT ACT_SITE 134 134 NUCLEOPHILE.
FT ACT_SITE 138 138 PROTON DONOR.
FT DISULFD 61 90
FT MUTAGEN 51 51 D->N: 30% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 89 89 D->N: 85% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 92 92 E->Q: 3% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 95 92 E->Q: 50% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 105 105 D->N: 15% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 133 133 E->Q: 0.2% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 134 134 D->N: 0.5% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 136 136 E->Q: COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 138 138 D->N: 65% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 143 143 E->Q: 15% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 160 160 D->N: 60% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 168 168 D->N: 80% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 179 179 D->N: 70% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 190 190 D->N: NO CHANGE IN ACTIVITY.
FT MUTAGEN 219 219
SQ SEQUENCE 243 AA; 27435 MW; 651188D9AAD609A5 CRC64;

Query Match 12.88; Score 233.5; DB 1; Length 243;
Best Local Similarity 33.7%; Pred. No. 7,7e-10;
Matches 58; Conservative 26; Mismatches 77; Indels 11; Gaps 6;
QY 23 TNVSAKDFSGAEIYLTLEVOYGFKEARMKMAASGVSSMFLYONGSEIADGRPWVEVDI 82

```

```

Db 81 TSPFNKFCGSGENSVOTYGYGLYEVNMKPAKNVGISSFFTYGPT---DGTWDELDI 137
QY 83 EVLGKNGSGSQSNITTKAGAKTSEKHNAVSPADAFFHTYGLEMTPNVRYVTVGOEV 142
Db 138 EFLKDKDTTKVQFNFTYGVG---NHEKIVNLGFDPAANSYHTYAFDQPNSTIKWVYDQ-L 193
QY 143 RKTGGGVSNLTGTGRLFNLMSSESA-AWVGOFDESKLPLFOFINWVKYK 193
Db 194 KHTATTOIPQTPCK--IMNLMNGAGVDMLGSYN-GVTPLSRSLHWRTK 242

RESULT 8
YGA6_YEAST STANDARD: PRT: 507 AA.
ID YG6_YEAST STANDARD: PRT: 507 AA.
AC P53301:
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOHETICAL 52.8 KDA PROTEIN IN BUB1-HIP1 INTERGENIC REGION.
YCR189C OR G7553.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=97279231; PubMed=9133739;
RA Attyo J., Garcia-Gonzalez M., Garcia-Saez M.I., Sanchez-Perez M.,
RA Nombela C.;
RT "DNA sequence analysis of a 23,002 bp DNA fragment of the right arm
RT of Saccharomyces cerevisiae chromosome VII."
RL Yeast 13:357-363(1997).
CC -1- SIMILARITY: SOME, TO YEAST UTR2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 272974; CAA97215.1; -
DR EMBL: X99074; CAA67525.1; -
DR HSSP: P23904; IAUO.
DR SGD: S0003421; CRH1.
DR InterPro: IPR000757; -
DR Pfam: PF00722; Glyco_hydro_16; 1.
KM Hypothetical protein.
FT DOMAIN 63 66 POLY-SER.
FT DOMAIN 301 310 POLY-SER.
FT DOMAIN 345 357 POLY-SER.
FT DOMAIN 387 391 POLY-SER.
FT DOMAIN 467 470 POLY-SER.
FT DOMAIN 507 52757 MM; 7D7B61F57AE942C CRC64;
SQ SEQUENCE 507 AA; 52757 MM; 7D7B61F57AE942C CRC64;

Query Match 12.6%; Score 230; DB 1; Length 507;
Best Local Similarity 26.9%; Pred. No. 3.3e-09;
Matches 87; Conservative 57; Mismatches 134; Indels 46; Gaps 13;

```

```

Db 180 LDWAMDKTWTYLDGESVYR-----VLSTNTSEGVPSQPMYLMGICAGDPDPAAGTIEW 233
QY 172 V-GQFDESKLPLFOFINWVKYKTPQGGSGDFTLDMWDNDFDTGSRMGKGDWTFDG 230
Db 234 AGGETNINDAPFTWYIEKTVITVDTSTGKRYTYGDQGSW-ESTLEADGGSYGYDAQE- 291
QY 231 NRVDLTGKNINYSRGMILALTRKG--QESFNGQVPRDEPAQSSSAPASS--SSVPA 286
Db 292 -----DFVALANGSLSSSTSTSSSTVSSASSSTVSSSVSSSTVSSSVSSSTVS 345
QY 287 SSSSVPASSSSAFVPPSSSATNA 310
Db 346 SSSSVSSSSSTSS--PSSSTATVS 366

RESULT 9
GUB_PAEMA STANDARD: PRT: 237 AA.
ID GUB_PAEMA STANDARD: PRT: 237 AA.
AC P23904;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
OS Paenibacillus macerans (Bacillus macerans).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Paenibacillus.
OX NCBI_TaxID=44252;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9109712; PubMed=2274030;
RA Borriss R., Buettner K., Maentzsaetae P.;
RT "Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans:
RT homologues to other beta-glucanases."
RL Mol. Gen. Genet. 222:278-283(1990).
RN [2]
RP ACTIVE SITE.
RX MEDLINE=93094208; PubMed=1360982;
RA Hoeft P.B., Condron R., Traeger J.C., McAniff J.C., Stone B.A.;
RT "Identification of glutamic acid 105 at the active site of Bacillus
RT amyloliquefaciens 1,3-1,4-beta-D-glucan 4-glucanohydrolase using
RT epoxide-based inhibitors."
RL J. Biol. Chem. 267:25059-25066(1992).
RN [3]
RP MUTAGENESIS OF GLU-128.
RA Olsen O.;
RL Thesis (1990), University of Aarhus, Denmark.
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=93281743; PubMed=8099449;
RA Ketel T., Simon O., Borriss R., Helmann U.;
RT "Molecular and active-site structure of a Bacillus 1,3-1,4-beta-
RT glucanase."
RL Proc. Natl. Acad. Sci. U.S.A. 90:5287-5291(1993).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RX MEDLINE=96028129; PubMed=7588726;
RA Hahn M., Ketel T., Helmann U.;
RT "Crystal and molecular structure at 0.16-nm resolution of the hybrid
RT Bacillus endo-1,3-1,4-beta-D-glucan 4-glucanohydrolase H(A16-M)."
RL Eur. J. Biochem. 232:849-858(1995).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```



```

Qy 28 KDFSGAEIYLTLEEVYOGCFEARMKMAAGSVSMFLYONGSEIADGRPWVEVDIEVLGK 87
Db 3 KNSGDTVLSSTRAWYMGKVASRIKSHLAGVYTGITLISGAGD-----ELDPEFGA 54
Qy 88 NPGSQSNI-----ITGKAGAOQTSKHNHAVSPADQAFHTYGLFVYVYKMTVDQ 140
Db 55 DLEFQOTNFYVESVLYNTNSANISPTD-----TFENYHYELDHDHEDVVTMSIDGV 105
Qy 141 EVRTEGGQVGNLNGTOGLRF-----NLW-----SSESA-----AMWG-----QPEESKLP 181
Db 106 VGRILYKMEYTNAT-TQKYQYFQPIPSKVDISITWPGCNSTNAPGTIAMSGETIHWASDIS 164
Qy 182 ----LFOFINWVKVYKYPGOGEGSDYTLDTWDTNFDPFDSRMGKGWTFDGNHVDLTD 237
Db 165 NPGYYAIVNEVNITCYP-----PSDIKKNGTSAVYVTSSE-----FLAKIATLD 212
Qy 238 KNIV--SFDGMILALTRKKGSEFNGQVPRDEPAPQSSSAPASSSVPAS----- 287
Db 213 DEVAMDSEEG-----SGLDPHKGAT-----TSSTQKSSSTATSSKTSDDHSSSTKK 260
Qy 288 ---SSSVAPASSSAFVPPSSSSATTAHGMKTPAVA 321
Db 261 SSKTSSTASSSSSSSSSSSSSSSTATKNGDKVSVS 297

RESULT 13
MERS_ARATH
AC P24806;
STANDARD; PRT; 247 AA.
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE MERI-5 PROTEIN.
GN MERI-5.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
NCBI-TaxID=3702;
RX MEDLINE=93005704; PubMed=1840916;
RA Medford J.I., Elmer J.S., Klee H.J.;
RT "Molecular cloning and characterization of genes expressed in shoot
RT apical meristems.";
RL Plant Cell 3:359-370(1991).
CC -I- TISSUE SPECIFICITY: EXPRESSED IN SHOOT APICAL MERISTEMS, ALSO
CC FOUND IN SEEDLINGS AND MERISTEMS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63166; AAA32828.1;
DR PIR; J01022; J01022.
DR HSSP; P23904; IAKK.
DR InterPro; IPR000757;
DR Pfam; Pf00722; Glyco_hydro_16; 1
DR SEQUENCE 247 AA; 28295 MW; 831E8441564B45E8 CRC64;

Query Match 8.4%; Score 154; DB 1; Length 247;
Best Local Similarity 26.5%; Pred. No. 0.00032;
Matches 83; Conservative 49; Mismatches 120; Indels 85; Gaps 15;

```

	Matches	53;	Conservative	34;	Mismatches	71;	Indels	42;	Gaps	9
QY	6	TAKVSALAAVAAAAAALLTTNVS-----AKDFSGAEIYLTE-----EVOYGKFE	47							
Dd	11	TLTVAASAVSADNTPTVTYNVAMNGRGRLINNLGGTLTDLSKSSGSGFSQSTXEYLFGRKD	70							
OY	48	ARMMAA--ASGTATSMPELYONGSEIADGRPMVEVDIEVLK---NPGSFQSNITTKAG	102							
Dd	71	MOIKLVGNNSAGIVTYTRFLKSEG-----TWDEIDFELGMSSDPTLTHTNVYTQSK	124							
QY	103	AOKTSEKHAAVSPAADQAFTHTYGLEMTPNVRYMTVDG--EVRKTGGGVSNLTCGLR	160							
Dd	125	-DKROQHLEMDPTAN--FHHTYSTILMNQRILLFYDTHPLSLKTMSLYSVSCFOFKTKPBR	181							
QY	161	F-----NLMSESAAW	172							
Dd	182	MYROFMEROSDMATKRRSWL	201							
RESULT	14									
ID	GOB_RHOMR	STANDARD;	PRT;	286	AA.					
AC	P45798;									
DT	01-NOV-1995 (Rel. 32, Created)									
DT	01-NOV-1995 (Rel. 32, Last sequence update)									
DE	BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)									
DN	(1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).									
GN	BGLA.									
OS	Rhodothermus marinus.									
OC	Bacteria; CFB group; Rhodothermus group; Rhodothermus.									
OX	NCBI_TaxID=29549;									
RN	[1]									
RP	SEQUENCE FROM N.A.									
RC	STRAIN=21 / ITI-378;									
RX	MEDLINE=95010084; Pubmed=7925416;									
RA	Splittaeet R., Hreggvidsson G.O., Kristjansson J.K.,									
RT	Eggerudsson G., Palsdottir A.;									
RF	"Cloning and sequencing of a Rhodothermus marinus gene, bglA, coding									
RT	for a thermostable beta-glucanase and its expression in Escherichia									
coll."										
CC	Eur. J. Biochem. 224:923-930(1994).									
CC	-!- FUNCTION: SHOWS ACTIVITY ON LICHENARY BETA-GLUCAN AND LAMINARIN									
CC	BUT NOT ON CMC CELLULOSE OR XYLAN. THE ENZYME HAS A TEMPERATURE									
CC	OPTIMUM OF 85 DEGREES CELSIUS AND A PH OPTIMUM OF 7.0.									
CC	-!- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC									
CC	LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.									
CC	-!- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.									
CC	-----									
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration									
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -									
CC	the European Bioinformatics Institute. There are no restrictions on ways									
CC	use by non-profit institutions as long as its content is in no way									
CC	modified and this statement is not removed. Usage by and for commercial									
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/									
CC	or send an email to license@isb-sib.ch).									
CC	-----									
DR	EMBL; U04836; AAA60459.1; --									
DR	HSSP; P23904; IAJK.									
DR	InterPro; IPR000757; -									
DR	Pfam; PF00722; Glyco_hydro_16.1.									
DR	PROSITE; PS01034; GLYCOSYL_HDROL_F16; 1.									
KM	Hydrolyase; Glycosidase; Signal.									
FT	CHAIN	1	30	POTENTIAL.						
FT	ACT_SITE	158	158	BETA-GLUCANASE.						
FT	ACT_SITE	163	163	NUCLEOPHILE (BY SIMILARITY).						
FT	ACT_SITE	286	AA;	PROTON DONOR (BY SIMILARITY).						
SQ	SEQUENCE	286	AA;	33145 MW; 7215C33624135191 CRC64;						
Query Match	Best Local Similarity	23.7%	Pred. No. 0.011;	Score 133.5; DB 1; Length 286;						

Matches 250: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 2 VSADFGAEIYLTLEEVYQGFKEARMKMAAAGTSSMFLYONGSEIADGRPWVEVDIEV 61
 |||||
 Db 25 VSADFGAEIYLTLEEVYQGFKEARMKMAAAGTSSMFLYONGSEIADGRPWVEVDIEV 84
 |||||
 QY 62 LGKNGSFGOSNIITGKAGAQKTSKHHAVSPADDAEFTYCLEWTPNYVWTVDGQEVK 121
 |||||
 Db 85 LGKNGSFGOSNIITGKAGAQKTSKHHAVSPADDAEFTYCLEWTPNYVWTVDGQEVK 144
 |||||
 QY 122 TEGGVSNTLGTGTLRNLMSSESAAMVQGFDESKLPFQFINWVKYKKTGCGEGGSD 181
 |||||
 Db 145 TEGGVSNTLGTGTLRNLMSSESAAMVQGFDESKLPFQFINWVKYKKTGCGEGGSD 204
 |||||
 QY 182 FTLDWTDNFTFDGSRNGKDMFTDGNRVDLTDKNYSRQGLILALTRKQESFNGCVP 241
 |||||
 Db 205 FTLDWTDNFTFDGSRNGKDMFTDGNRVDLTDKNYSRQGLILALTRKQESFNGCVP 264
 |||||
 QY 242 RDDEPAPNSSS 252
 |||||
 Db 265 RDDEPAPNSSS 275

RESULT 2

GUB_BACBR STANDARD; PRT: 259 AA.

ID GUB_BACBR

AC P37073;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last annotation update)

DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)

DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).

GN Bg1AB.

OS Bacillus brevis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Brevibacillus.

OX NCBI_TaxID=1393;

RA MEDLINE=93159752; PubMed=7763386;

RA Louw M.E., Reid S.J., Watson T.G.;

RT "Characterization, cloning and sequencing of a thermostable endo-(1,3-1,4) beta-glucanase-encoding gene from an alkalophilic Bacillus brevis";

RL Appl. Microbiol. Biotechnol. 38:507-513(1993).

CC -1- FUNCTION: HYDROLYSES B-GLUCANS CONTAINING MIXED BETA-1,3 AND BETA-1,4 LINKAGES. THIS ENZYME IS THERMOSTABLE; ITS OPTIMAL TEMPERATURE IS NEAR 65-70 DEGREES CELSIUS.

CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.

CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE SIMILAR TO LICHENASE OF GERMINATING BARLEY.

CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: M84339; AAA22265.1; -

CC HSSP: P23904; IAKK.

DR InterPro: IPR000757; -

DR Pfam: PF00722; Glyco_hydro_16; 1.

DR PRINTS: PR00737; GLHYDRLASE16.

DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.

KW Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 31

FT CHAIN 32 259

FT ACT_SITE 142 146

FT ACT_SITE 146 146

PROTON DONOR (BY SIMILARITY).

SQ SEQUENCE 259 AA; 29960 MW; A63C09F281FF5D13 CRC64;

Query Match 18.5%; Score 266.5; DB 1; Length 259;

Best Local Similarity 36.08; Pred. No. 3e-15;

Matches 63: Conservative 24; Mismatches 75; Indels 13; Gaps 6;

QY 3 SAKDFSGAEIYLTLEEVYQGFKEARMKMAAAGTSSMFLYONGSEIADGRPWVEVDIEV 62
 |||||
 Db 90 SARYKAGELRTNDFYHYGLFEVSMKPAKVEGYTSSEFTY-TGEMDKDGDPMDEIDIEFL 148
 |||||
 QY 63 GRNPGSFGOSNIITGKAGAQKTSKHHAVSPADDAEFTYCLEWTPNYVWTVDGQEVKRT 122
 |||||
 Db 149 GKDTTRIQFNYFTNGVCG---NEFYDLDGFDASESFTYAFEREDSITWYVNGEAVHTA 205
 |||||
 QY 123 EGGGVSNLTGTLRNLMSSESA-AMVQGFDESKLPFQFINWVKYKKTGCGEGGSD 175
 |||||
 Db 206 ---TENIPQTPKIMNLMFVGVDGTVFDGDNTPVSYDWW---RYTPIQ 253
 |||||

RESULT 3

GUB_BACSU STANDARD; PRT: 242 AA.

ID GUB_BACSU

AC P04957;

DT 13-AUG-1987 (Rel. 05, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)

DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).

GN BGLS OR BGL OR LIC5 OR N15B.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1423;

RA MEDLINE=84272222; PubMed=6087283;

RA Murphy N., McConnell D.J., Cantwell B.A.;

RT "The DNA sequence of the gene and genetic control sites for the excreted B. subtilis enzyme beta-glucanase";

RL Nucleic Acids Res. 12:5355-5367(1984).

RN [2]

RN SEQUENCE FROM N.A.

RC STRAIN-HL-25;

RA Tezuka H., Yuuki T., Yabuuchi S.;

RT "Construction of a beta-glucanase hyperproducing Bacillus subtilis using the cloned beta-glucanase gene and a multi-copy plasmid";

RL Agric. Biol. Chem. 53:2335-2339(1989).

RN [3]

RN SEQUENCE FROM N.A.

RA STRAIN-168 / BGSJAL;

RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;

RT Submitted (FEB-1995) to the EMBL/GenBank/DBD databases.

RL [4]

RN SEQUENCE FROM N.A.

RC STRAIN-168;

RX MEDLINE=95219081; PubMed=7704256;

RA Wolf M., Gezei A., Simon O., Borries R.;

RT "Genes encoding xylan and beta-glucan hydrolasing enzymes in Bacillus subtilis: Characterization, mapping and construction of strains deficient in lichenase, cellulase and xylanase";

RL Microbiology 141:281-290(1995).

RN [5]

RN SEQUENCE OF 1-6 FROM N.A.

RC STRAIN-BR151;

RX MEDLINE=96178961; PubMed=8606172;

RA Schneitz K., Stuelke J., Gertz S., Krueger S., Krieg M., Hecker M., Rak B.;

RT "Bact, a Bacillus subtilis transcriptional antiterminator protein of the BglG family";

RL J. Bacteriol. 178:1971-1979(1996).

RN [6]

```

DE (EC 3.2.1.8) (XYLANASE); (ENNO-BETA-1,3-1,4 GLUCANASE) (EC 3.2.1.73)
DN 1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE)].
OS Ruminococcus flavefaciens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Ruminococcus.
OC NCBI_TaxID=1265;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17;
RX MEDLINE=93259938; PubMed=8491715;
RA Flint H.J., Martin J., McPherson C.A., Daniel A.S., Zhang J.-X.;
RT "A bifunctional enzyme, with separate xylanase and beta(1,3-1,4)-
RT glucanase domains, encoded by the xynD gene of Ruminococcus
RT flavefaciens."
RL J. Bacteriol. 175:2943-2951(1993).
CC -1- FUNCTION: CONTAINS TWO CATALYTIC DOMAINS WITH XYLANASE AND ENDO-
CC BETA-1,3-1,4 GLUCANASE ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: ENDOTHEROLYSIS OF 1,4-BETA-D-XYLOSIDIC
CC LINKAGES IN XYLANS.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- PATHWAY: XYLAN DEGRADATION.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC G (FAMILY 11 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 16 OF
CC GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; S61204; AAB26620.1; -.
DR HSSP; P23904; IAJK.
DR InterPro; IPR000757; -.
DR InterPro; IPR001137; -.
DR Pfam; PF02018; CBD_6; 1.
DR Pfam; PF00457; Glyco_hydro_11; 1.
DR Pfam; PF00722; Glyco_hydro_16; 1.
DR PRINTS; PR00737; GLYHIDRLASE16.
DR PRINTS; PR00911; GLHYDRASE11.
DR PROSITE; PS00776; GLYCOSYL_HYDROL_F11_1; 1.
DR PROSITE; PS00777; GLYCOSYL_HYDROL_F11_2; 1.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
DR KMW Xylan degradation; Hydrolyase; Glycosidase; signal;
KW Multifunctional enzyme.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 802 XYLANASE/BETA-GLUCANASE.
FT DOMAIN 32 244 A (XYLANASE).
FT DOMAIN 245 523 B.
FT DOMAIN 524 555 LINKER.
FT DOMAIN 556 802 C (BETA-GLUCANASE).
FT ACT_SITE 124 124 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 226 226 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 684 684 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 524 529 POLY-THR.
FT DOMAIN 532 543 POLY-THR.
FT DOMAIN 546 553 POLY-THR.
SQ SEQUENCE 802 AA; 89091 MW; 2880A689E447284AF CRC64;

Query Match 16.9%, Score 242.5; DB 1; Length 802;
Best Local Similarity 34.9%; Pred. No. 1.le-12;
Matches 58; Conservative 21; Mismatches 74; Indels 13; Gaps 6;

7 FSGAEIYLFLEEVCKFEARKMMAASGVSSMELFYONGSEIADGRPAPVEDIEVLGNP 66
DB 634 YSGGEFRNNFYHYGYEPCSMQAMKNDGVSSFEFTYTGPS--DDNPWDEIDIEILGKNT 690

```

QY 67 GSFQSNITGKAGAKTSEKHAAVSPAQAQFHTYGLEMTPNVYRWTVDOEVRKTEGQ 126
 DB 691 TOYQFNNTYGGCKH---EKYLDLGFDSSEAHYHTYGFDMQPNYIAWVDREYRA---- 743
 QY 127 VSNITGOG-LRFNLSSESA-AMVGFDESKLPLFOFIMWVYK 170
 DB 744 TODIPKTPKIMNAMPGLTVDMKLAFN-GRTPLTAAHYOMVYK 788

RESULT 5

GUB_BACAM STANDARD: PRT; 239 AA.
 ID P07980;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
 GN BGLA.
 OS *Bacillus amyloliquefaciens*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1390.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BE 20/78;
 RX MEDLINE=87192007; PubMed=3106158;
 RA Hofmeister J., Kurtz A., Borris R., Knowles J.;
 RT "The beta-glucanase gene from *Bacillus amyloliquefaciens* shows
 extensive homology with that of *Bacillus subtilis*.";
 RL Gene 49:177-187(1986).
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
 CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
 CC MISCELLANEOUS: BETA-GLUCANASES OF *BACILLUS* HAVE A SUBSTRATE RANGE
 CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M15674; AAA87323.1;
 CC DR PIR: A29091; A29091.
 CC DR HSSP: P27051; 1GBG.
 CC DR InterPro: IPR000757;
 CC DR Pfam: PF00722; GLYCO_hydro.16; 1.
 CC DR PRINTS: PR00737; GLHYDRLASE16.
 CC DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
 CC KM Hydrolyase; Glycosidase; Signal.
 CC FT SIGNAL 1 25
 CC FT CHAIN 26 239 BETA-GLUCANASE.
 CC FT ACT_SITE 134 86 NUCLEOPHILE (BY SIMILARITY).
 CC FT DISULFID 57 86 BY SIMILARITY.
 CC SQ SEQUENCE 239 AA; 26928 MW; A76A64268A7AA0B CRC64;

Query Match 16.6%; Score 238.5; DB 1; Length 239;
 Best Local Similarity 34.3%; Pred. No. 5.7e-13;
 Matches 58; Conservative 26; Mismatches 74; Indels 11; Gaps 6;

QY 3 SAKDFGAEIYLTLEVOYKFEARMMAASGVSSNFTYONGSEIADGRWVEVDIEVL 62
 DB 80 SYNKFCCGGERSVQYTGYGIVEMRKPAKNTGIVSSFFTYTGPT--GTPMDEIDIEFL 136
 QY 63 GRNPGSFQSNITGKAGAKTSEKHAAVSPAQAQFHTYGLEMTPNVYRWTVDOEVRK 122
 DB 137 GKTTTVOQFNNTYGGCKH---NHKEPADLGDANANAHYTAEDQPNYSIKWYVGO-LKRT 192
 QY 123 EGGQVSNLTGTOGLRPNLSSESA-AMVGFDESKLPLFOFIMWVYK 170

DB 193 ATTQIPAPK--IMNMIMGTCVDDMLGSGYCNV-PIYAHYDMRYRK 238

RESULT 6

GUB_CLOTM STANDARD: PRT; 334 AA.
 ID P29716; P37074;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LAMINARINASE).
 GN LICB OR LAM1.
 OS *Clostridium thermocellum*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1515;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 27405 / DSM 1237;
 RX MEDLINE=9255194; PubMed=1740123;
 RA Schimming S., Schwarz W.H., Staudenbauer W.L.;
 RT "Structure of the *Clostridium thermocellum* gene *licB* and the encoded
 RT beta-1,3-1,4-glucanase. A catalytic region homologous to *Bacillus*
 RT cellulasenes joined to the reiterated domain of clostridial
 RT cellulases.";
 RL Eur. J. Biochem. 204:13-19(1992).
 RN [2]
 RP PRELIMINARY SEQUENCE FROM N.A.
 RC STRAIN-E7;
 RX MEDLINE=92095946; PubMed=1755832;
 RA Zverlov V.V., Laptev D.A., Tishkov V.I., Velikodvorskaja G.A.;
 RT "Nucleotide sequence of the *Clostridium thermocellum* laminarinase
 RT gene.";
 RL Biochem. Biophys. Res. Commun. 181:507-512(1991).
 RN [3]
 RP REVISIONS.
 RA Zverlov V.V.;
 RL Submitted (Aug-1993) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP CHARACTERIZATION.
 RA Zverlov V.V., Velikodvorskaja G.A.;
 RC STRAIN-E7;
 RT "Cloning the *Clostridium thermocellum* thermostable laminarinase gene
 RT in *Escherichia coli*; the properties of the enzyme thus produced.";
 RL Biotechnol. Lett. 12:811-816(1990).
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
 CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
 CC -1- SUBUNIT: MAY FORM PART OF A MULTISUBUNIT COMPLEX (CELLULOSE).
 CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
 CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMS. THIS DOMAIN
 CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X63355; CAA4959.1;
 CC DR EMBL: X58392; CAA41281.1;
 CC DR PIR: S23498; S23498.
 CC DR PIR: J50611; J50611.
 CC DR PIR: S18726; S18726.
 CC DR HSSP: P23904; 1AJK.
 CC DR InterPro: IPR000757;
 CC DR InterPro: IPR002048;
 CC DR InterPro: IPR002105;
 CC DR Pfam: PF00404; Dockerin_1; 2.

```

DR Pfam; PF00722; Glyco_hydro_16; 1.
DR PRINTS; PR00737; GLHYDRLASE16.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
DR PROSITE; PS00018; EF_HAND-UNKNOWN_2; 1.
DR PROSITE; PS00448; CLOS_CELULOSOME_RPT; 2.
KW Hydrolyase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 27
FT CHAIN 28 334
FT ACT_SITE 136 136 BETA-GLUCANASE.
FT ACT_SITE 140 140 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 252 269 PROTON DONOR (BY SIMILARITY).
FT DOMAIN 273 331 PRO/THR-RICH (LINKER).
FT REPEAT 273 296 2 x 24 AA APPROXIMATE REPEATS.
FT REPEAT 308 331 1.
FT REPEAT 308 331 2.
FT CONFLICT 304 334 OSVADVNRDRGRIDSTDLMLKRYLIRAIPLSL ->
FT SEQUENCE 334 AA; 37897 MW; 0837564E9726F281 CRC64;
PODCGGRDVRVDSGSK (IN REF. 2 AND 3).

Query Match 16.5%; Score 237; DB 1; Length 334;
Best Local Similarity 30.9%; Pred. No. 1,le-12;
Matches 80; Conservative 27; Mismatches 94; Indels 58; Gaps 13;

QY 5 KDFSGAELTYLEVO-----YGFKFAKMAAASGVSSMFLYONGSEIADGRPWEVD 58
DB 82 REYGSGYKSGEYRTKSGFGYGEYVRMAKAKNGIVSFEFTYGPS---DNNPWEID 138
QY 59 IEVIGKNGSQSNITITKAGAKTSEKHNAVSPADAFTHTYGLEMTPNVYRWTVDOGE 118
DB 139 IEFLKDTKTVQVPMWYKNGVSG---NEYLHMGFDASODFHTYGEWMPDYIDFYVDGK 195
QY 119 VAKTEGGGVSNLTGTQG-LRFNLMSESA-AWVGQFDESKLPLFQFIWVKY----- 169
DB 196 VYR-----GTRNIPVTPGKIMNMLWPGIGVDEVLGRYD-GRIPLQAEYEVKTYRPGVPOD 250
QY 170 -----KTFP-----GQGE-GGSDFT-----LDWTNDFPFDSRWKSG 201
DB 251 NPTPTPTAPSTPTPNPLKGDVNGDGHVNSDYSLEKRYLIRVYIDREFVDDGS---VA 307
QY 202 DMTFPGNRVDLTDKNIYSR 220
DB 308 DVNRDG-RIDSTDLTMLKR 325

RESULT 7
GUB_BACLI STANDARD: PRT: 243 AA.
ID GUB_BACLI
AC P27051;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
GN Bgl.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OC NCBI_TaxID=1402;
RX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91224124; PubMed=2026156;
RA Lloberas J., Perez-Pons J.A., Querol E.;
RT "Molecular cloning, expression and nucleotide sequence of the
RT endo-beta-1,3-1,4-D-glucanase gene from Bacillus licheniformis.
RT Predictive structural analyses of the encoded polypeptide.";
RL Eur. J. Biochem. 197;337-343(1991).
RN [2]
RP REVISIONS.
RA Querol E.;
RN Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
RX [3]
RX MUTAGENESIS.
RX MEDLINE=92362869; PubMed=1354172;

```

```

RA Planas A., Juncosa M., Lloberas J., Querol E.;
RT "Essential catalytic role of Glu134 in endo-beta-1,3-1,4-D-glucan 4-
RT glucanohydrolase from B. licheniformis as determined by site-directed
RT mutagenesis.";
RL FEBS Lett. 308;141-145(1992).
RN [4]
RP MUTAGENESIS.
RX MEDLINE=94237863; PubMed=8182059;
RA Juncosa M., Pons J., Dot T., Querol E., Planas A.;
RT "Identification of active site carboxylic residues in Bacillus
RT licheniformis 1,3-1,4-beta-D-glucan 4-glucanohydrolase by
RT site-directed mutagenesis.";
RL J. Biol. Chem. 269;14530-14535(1994).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=96063718; PubMed=7589539;
RA Hahn M., Pons J., Planas A., Querol E., Heinemann U.;
RT "Crystal structure of Bacillus licheniformis 1,3-1,4-beta-D-glucan 4-
RT glucanohydrolase at 1.8-A resolution.";
RL FEBS Lett. 374;221-224(1995).
CC -I- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -I- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
CC -I- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X57279; CA940547.1; -.
DR PIR; S15368; S15368.
DR PDB; 1GBG; 07-DEC-95.
DR InterPro; IPR000757; -.
DR Pfam; PF00722; Glyco_hydro_16; 1.
DR PRINTS; PR00737; GLHYDRLASE16.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
KW Hydrolyase; Glycosidase; Signal; 3D-structure.
FT SIGNAL 1 27
FT CHAIN 28 243
FT ACT_SITE 134 134 BETA-GLUCANASE.
FT ACT_SITE 138 138 NUCLEOPHILE.
FT DISULFID 61 90 PROTON DONOR.
FT MUTAGEN 51 51
FT MUTAGEN 89 89 D->N: 30% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 92 92 D->N: 85% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 95 95 E->Q: 3% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 105 105 E->Q: 50% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 133 133 D->N: 15% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 134 134 D->N: 0.2% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 136 136 D->N: 0.5% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 138 138 E->Q: COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 143 143 D->N: 65% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 160 160 E->Q: 15% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 168 168 D->N: 60% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 179 179 D->N: 80% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 190 190 D->N: 70% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 219 219 D->N: NO CHANGE IN ACTIVITY.
FT SEQUENCE 243 AA; 27435 MW; 651188D9AAD609A5 CRC64;

Query Match 15.9%; Score 229.5; DB 1; Length 243;
Best Local Similarity 33.7%; Pred. No. 3,3e-12;
Matches 57; Conservative 25; Mismatches 76; Indels 11; Gaps 6;

QY 3 SAKDFSGAELTYLEVOYGFKFAKMAAASGVSSMFLYONGSEIADGRPWEVDIEVL 62
DB 84 SYNFKDGENSVQTYGYGLETVNKKPAKNGIVSFEFTYGP1---DGTWDEIDIEFL 140
QY 63 GKNPDSFOSNITITKAGAKTSEKHNAVSPADAFTHTYGLEMTPNVYRWTVDOGEVAKT 122

```

DB 141 GKDTKVOFNYYTNGV---NHEKIVNLGFDANSHYHTAFDMQPSIKWYVDCO-LKHT 196
 OY 123 EGGVSNLTGNGGLRFNLMSESA-AWVGOPDESKLPFOFINWVYK 170
 DB 197 ATTQIPOTPGK--IMNMLMGAGVDENLGSYN-GVTPLSRSLHWRYTK 242

RESULT 8
 GUB_PAEMA STANDARD; PRT; 237 AA.
 ID GUB_PAEMA
 AC P23904;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 OS (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE)
 OC Paenibacillus macerans (Bacillus macerans)
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Paenibacillus.
 OX NCBI_TaxID=44252;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91109712; PubMed=2274030;
 RA Borriss R., Buettner K., Maentzsaellae P.;
 RT "Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans:
 RT homologues to other beta-glucanases.";
 RL Mol. Gen. Genet. 222:278-283(1990).
 RN [2]
 RP ACTIVE SITE
 RX MEDLINE=93094208; PubMed=1360982;
 RA Hoel P.B., Condon R., Traeger J.C., McAniff J.C., Stone B.A.;
 RT "Identification of glutamic acid 105 at the active site of Bacillus
 RT amyloliquefaciens 1,3-1,4-beta-D-glucan 4-glucanohydrolase using
 RT epoxide-based inhibitors.";
 RL J. Biol. Chem. 267:25059-25066(1992).
 RN [3]
 RP MUTAGENESIS OF GLU-128.
 RA Olsen O.;
 RL Thesis (1990), University of Aarhus, Denmark.
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=93281743; PubMed=8099449;
 RA Ketel T., Simon O., Borriss R., Heinemann U.;
 RT "Molecular and active-site structure of a Bacillus 1,3-1,4-beta-
 RT glucanase".
 RL Proc. Natl. Acad. Sci. U.S.A. 90:5287-5291(1993).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RX MEDLINE=96028129; PubMed=7588726;
 RA Hahn M., Ketel T., Heinemann U.;
 RT "Crystal and molecular structure at 0.16-nm resolution of the hybrid
 RT Bacillus endo-1,3-1,4-beta-D-glucan 4-glucanohydrolase H(A16-W)".
 RL Eur. J. Biochem. 232:849-858(1995).
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
 CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
 CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
 CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: X55959; CAA39426.1; -
 CC PIR: S11927; S11927.
 DR PDB: 2AYH; 31-MAR-95.
 DR PDB: 1BYH; 31-OCT-93.
 DR PDB: 1CPM; 22-JUN-94.

DR PDB: 1CPN; 22-JUN-94.
 DR PDB: 1MAC; 27-FEB-95.
 DR PDB: 1AJK; 06-MAY-98.
 DR PDB: 1AJO; 06-MAY-98.
 DR InterPro: IPR000757;
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 DR PRINTS; PR00737; GLHYDRLASE16.
 DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
 DR KMW Hydrolyase; Glycosidase; Signal; 3d-structure.
 FT SIGNAL 1 23
 FT CHAIN 24 237
 FT ACT_SITE 128 128
 FT ACT_SITE 132 132
 FT DISULFID 55 84
 FT NOTAGEN 128 128
 FT STRAND 26 26
 FT STRAND 29 31
 FT STRAND 41 43
 FT TURN 51 52
 FT STRAND 53 53
 FT STRAND 56 57
 FT STRAND 59 61
 FT HELIX 62 64
 FT STRAND 66 67
 FT TURN 70 78
 FT TURN 79 80
 FT STRAND 81 89
 FT STRAND 93 93
 FT STRAND 96 103
 FT TURN 108 109
 FT STRAND 110 118
 FT HELIX 120 122
 FT TURN 123 123
 FT STRAND 127 134
 FT TURN 135 136
 FT STRAND 137 147
 FT TURN 148 149
 FT STRAND 150 150
 FT STRAND 155 158
 FT TURN 163 164
 FT STRAND 168 174
 FT STRAND 179 183
 FT TURN 184 185
 FT STRAND 186 191
 FT STRAND 200 209
 FT TURN 213 216
 FT STRAND 225 236
 SO SEQUENCE 237 AA; 26589 MW; 436EABCDFFC87781 CRC64;

Query Match 15.7%; Score 226; DB 1; Length 237;
 Best Local Similarity 35.1%; Pred. No. 6.2e-12;
 Matches 59; Conservative 23; Mismatches 70; Indels 16; Gaps 8;

OY 7 FSGAEIYTLLEVOYGFREARKMAAAGTVSSMFLYQNGSEIADGRPWVEVDIEVLKNP 66
 DB 82 FDCAEKSTNIYGYGLYEVKMPAKNTGIVSFFYTGP---AHGTQWDEIDIEFLGKDT 138
 OY 67 GSGQSNITTKAKAQTSEKHNHVAAPADAFTTYGLETPNTVKNRTVDCQEVKRTBGG 126
 DB 139 TKQFNYNYTNGVGH---ELVISLGFDAKSGFHTYAFDMQPGYIKWYVDG-VLKHT---A 191
 OY 127 VSNLTGNG-LRENLMSSESA-AWVGOPDESKLPFOFINWVYK 172
 DB 192 TANIPSTPGKIMNLMNGGVDMGLDGSYNGAN-PLTAEYDMV---KYT 235

RESULT 9
 GUB_PAEMA STANDARD; PRT; 238 AA.
 ID GUB_PAEMA
 AC P45797;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 GN (1,3-1,4-BETA-4-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
 OS Paenibacillus polymyxa (Bacillus polymyxa).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Paenibacillus.
 NCBI_TaxID=1406;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 842;
 RX MEDLINE=92041687; PubMed=1938968;
 RA Gosalbes M.U., Perez-Gonzalez J.A., Gonzalez R., Navarro A.;
 RT "Two beta-glucanase genes are clustered in Bacillus polymyxa:
 RT molecular cloning, expression, and sequence analysis of genes
 RT encoding a xylanase and an endo-beta-(1,3)-(1,4)-glucanase.";
 RL J. Bacteriol. 173:7705-7710(1991).
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
 CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
 CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
 CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X57094; CAA40379.1; -
 DR HSSP; P23904; ICPM.
 DR InterPro: IPR000757; -
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 DR PRINTS; PR00737; GLYHRLASE16.
 DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
 KW Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 26
 FT CHAIN 27 238
 FT ACT_SITE 129 129 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 133 133 PROTON DONOR (BY SIMILARITY).
 FT DISULFID 56 85
 FT SEQUENCE 238 AA; 26919 MW; C0CF7B4E5D40E8C CRC64;
 SO
 Query Match 15.6%; Score 225; DB 1; Length 238;
 Best Local Similarity 34.5%; Pred. No. 7.5e-12;
 Matches 58; Conservative 23; Mismatches 71; Indels 16; Gaps 8;
 Oy 7 FSGAEVLYLEVOYGFKFAKMAAASGVSSMFLYONGSEIADGRPVVEVDIEVLGNP 66
 Db 83 FDCGEYRTNNYGYGLYEVSKPAKNTIVSSFFTYTSPH---GTOWDEIDIEFLGKDT 139
 Oy 67 GSPFSNITTKRAGAKTSEKHNAVSPADQAFHTYGLTEFNRYWYVWDGQEVKRTGEGQ 126
 Db 140 TKYQFNNTYTVGVGH---EKIINIGFDASTSEFHTYAFPMQGYIKWYVDG-VLKHT---A 192
 Oy 127 VSNLTGTGOG-LRFNLMSSEA-AWVGQFDESKLPLOFINWKKYKKT 172
 Db 193 TTNIPTSPFGKIMMNMNGTGYDSWLSYNGAN-PLYAEYDMV---KYT 236
 RESULT 10
 EXOK_RHIME STANDARD; PRT; 269 AA.
 AC P33693;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE ENDO-1,3-1,4-BETA-GLUCANASE EXOK PRECURSOR (EC 3.2.1.-) (SUCCINOGLYCAN
 DE BIOSYNTHESIS PROTEIN EXOK).
 GN EXOK.

OS Rhizobium meliloti (Sinorhizobium meliloti).
 OG Plasmid pSym (megaplasmid 2).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Rhizobiaceae; Sinorhizobium.
 NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-1021;
 RX MEDLINE=94042869; PubMed=8226645;
 RA Glucksmann M.A., Reuber T.L., Walker G.C.;
 RT "Family of glycosyl transferases needed for the synthesis of
 RT succinoglycan by Rhizobium meliloti.";
 RL J. Bacteriol. 175:7033-7044(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RGCR2011 / S047;
 RX MEDLINE=93241147; PubMed=8479421;
 RA Becker A., Kleckmann A., Arnold W., Pehler A.;
 RT "Analysis of the Rhizobium meliloti exoH/exoK/exoL fragment: ExoK
 RT shows homology to excreted endo-beta-1,3-1,4-glucanases and ExoH
 RT resembles membrane proteins.";
 RL Mol. Gen. Genet. 238:145-154(1993).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98226741; PubMed=9560202;
 RA York G.M., Walker G.C.;
 RT "The Rhizobium meliloti ExoK and ExoH glycanases specifically
 RT depolymerize nascent succinoglycan chains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:4912-4917(1998).
 CC -1- FUNCTION: CLEAVES HIGH MOLECULAR WEIGHT SUCCINOGLYCAN TO YIELD LHM
 CC SUCCINOGLYCAN. DYNAMICALLY REGULATE THE MOLECULAR WEIGHT
 CC DISTRIBUTION OF SUCCINOGLYCAN BY CLEAVING NASCENT SUCCINOGLYCAN
 CC ONLY DURING A LIMITED PERIOD AFTER ITS SYNTHESIS, PERHAPS BEFORE
 CC IT UNDERGOES A TIME-DEPENDENT CHANGE IN ITS CONFORMATION OR
 CC AGGREGATION STATE.
 CC -1- PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L20758; AAA16048.1; -
 DR EMBL; Z17219; CAA78927.1; -
 DR HSSP; P23904; IAKJ.
 DR InterPro: IPR000757; -
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
 KW Exopolysaccharide synthesis; Glycosidase; Hydrolase; Signal; Plasmid.
 FT SIGNAL 1 27
 FT CHAIN 28 269
 FT CONFLICT 93 102 ACCEIOTRRK -> LRNPADQG (IN REF. 2).
 FT SEQUENCE 269 AA; 30083 MW; 073CC7ED65EF2611 CRC64;
 SO
 Query Match 13.7%; Score 197; DB 1; Length 269;
 Best Local Similarity 32.5%; Pred. No. 1.8e-09;
 Matches 50; Conservative 31; Mismatches 61; Indels 12; Gaps 6;
 Oy 2 VSADFSAELVLYLEVOYGFKFAKMAAASGVSSMFLYONGSEIADGRPVVEVDIEV 61
 Db 87 VKERNFACGEIQTFRKRFGYTEARIKADSGINSAFETYICP---ADKRPHEIDREY 143
 Oy 62 LKKNRGSQSN-IITGRKAGAKTSEKHNAVSPADQAFHTYGLTEFNRYWYVWDGQEV 120
 Db 144 LKNTAKVQINQYVSAKAGNEFLAD---VPGANGQGNDAFYWEKRIIRYVNGELV 199
 Oy 121 K-TEGGVSNLTGTGGLRFLNLSSESA-AWVGQF 152

RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
 RA Ayiles E., Bero A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
 RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
 RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
 RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
 RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
 RA Taylor P., Wei Y., Yelton M., Borstein D., Davis R.W.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B-6441;
 RX MEDLINE=94016558; PubMed=8411151;
 RA Melnick L., Sherman F.;
 RT "The gene clusters ARC and COR on chromosomes 5 and 10, respectively,
 RT of Saccharomyces cerevisiae share a common ancestry.";
 RL J. Mol. Biol. 233:372-388(1993).
 CC -1- SIMILARITY: SOME, TO YEAST YGR189C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, U18779; AAB65002.1; ALT_INT.
 DR EMBL, L22173; AAA34941.1;
 DR EMBL, S65964; AAD13975.1;
 DR EMBL, S66130; AAB28444.1;
 DR PIR, S30839; S30839.
 DR HSSP, P23904; IAO.
 DR SGG, S0000766; UTR2.
 DR InterPro: IPR000757;
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 FT DOMAIN 234 322
 FT SIGNAL 269 283
 FT CONFLICT 10 10 L->V (IN REF. 3).
 FT CONFLICT 171 171 A->R (IN REF. 3).
 FT CONFLICT 234 234 S->C (IN REF. 3).
 FT SEQUENCE 347 AA; 36692 MW; 1E8AFB862C4BB328 CRC64;
 QY
 Query Match 10.2%; Score 147; DB 1; Length 347;
 Best Local Similarity 25.1%; Pred. No. 3.6e-05;
 Matches 50; Conservative 30; Mismatches 71; Indels 48; Gaps 9;
 OY 5 KDFSGAELTYLEEVQ--YKFEARMKMAASGVSSMFLYONGSELDGRWVEVDIEVLK 64
 DB 3 KNSGTVLSTRVAVWGKVSARIKTSHLGAVTGFLYSAGD-----ELDEFEVGA 54
 OY 65 NPGSFOSNI-----ITGKAGAOQKTESEKHAIVSPAADAFTYGLTEPNTVYRWTVDO 117
 DB 55 DLEFQTNINWYMSVLYNTSANSISTD-----TFENYITFELDMEDYVTSIDV 105
 OY 118 EVRKTEGGQVSNLTGTLGIR-----NLW--SSESA---AWG---QFDESKLP 158
 DB 106 VGRITLYKNETYNAT--TKQYQYQPTSPKVDISIMPGNSTNAGTIAWSGCEINWDASDIS 164
 OY 159 ----LQFQINWKKVKKTP 173
 DB 165 NPGYVAIVNEVNTICYDP 183
 RESULT 14
 GUB_RHOMR STANDARD; PRT; 286 AA.
 ID GUB_RHOMR
 AC P45798;
 DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
 GN BGLA.
 OS Rhodothermus marinus.
 OC Bacteria; CF3 group; Rhodothermus group; Rhodothermus.
 OX NCBI_TaxId=29549;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-21 / ITI-378;
 RX MEDLINE=95010084; PubMed=7925416;
 RA Spilliaert R., Hreggvidsson G.O., Kristjansson J.K.,
 RA Eggertsson G., Palsdottir A.;
 RT "Cloning and sequencing of a Rhodothermus marinus gene, bglA, coding
 RT for a thermostable beta-glucanase and its expression in Escherichia
 RT coli.";
 RL Eur. J. Biochem. 224:923-930(1994).
 CC -1- FUNCTION: SHOWS ACTIVITY ON LICHENAN, BETA-GLUCAN AND LAMINARIN
 CC BUT NOT ON CMC, CELLULOSE OR XLAN. THE ENZYME HAS A TEMPERATURE
 CC OPTIMUM OF 85 DEGREES CELSIUS AND A PH OPTIMUM OF 7.0.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, U04836; AAA60459.1;
 DR HSSP, P23904; IAO.
 DR InterPro: IPR000757;
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
 KW Hydrolyase; Glycosidase; Signal.
 FT SIGNAL 1 30
 FT CHAIN 31 286
 FT ACT_SITE 158 158 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 163 163 PROTON DONOR (BY SIMILARITY).
 FT SEQUENCE 286 AA; 33145 MW; 7215C33624135191 CRC64;
 QY
 Query Match 9.3%; Score 133.5; DB 1; Length 286;
 Best Local Similarity 23.7%; Pred. No. 0.00037;
 Matches 44; Conservative 34; Mismatches 85; Indels 23; Gaps 8;
 OY 5 KDFSGAELTYLEEVQ--YKFEARMKMAASGVSSMFLYONGSELDGRWVEVDIEVDI 59
 DB 103 REYTSARLVTRGKASWTVYGRFELRLPSGRGTWPAIMLPDQYGYSAV--WDNGEIDI 161
 OY 60 -EVLGNP-----GSFOSNIITGKAGAOQKTESEKHAIVSPAADAFTYGLTEPNTVYRWTV 114
 DB 162 MEHVGFNPDPVAGHVTATKAYNLHGTQROGGSIR--VPAARDDFHYALAEWPEELRWTV 218
 OY 115 DGOEVKTEGGQVSNLTG-----TQGLRFNLWSSESAAMWQ--FDESKLPFQFINW 165
 DB 219 DDLVLYRFPNERLTDEADRWHPFDQPHLLIMNIAVGGAMGQGVDEAFPAQLLVVDY 278
 OY 166 VKVYKK 171
 DB 279 VRYRW 284
 RESULT 15
 BRU1_SOYBN STANDARD; PRT; 283 AA.
 ID BRU1_SOYBN
 AC P35694;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE BRASSINOSTEROID-REGULATED PROTEIN BRU1.
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 1;
 OC Fabales; Fabaceae; Papilionoideae; Glycine.
 ON NCBI_TaxID=3847;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Epicotyl:
 RX MEDLINE=84159788; PubMed=8115544;
 RA Zurek D.M., Clouse S.D.;
 RT "Molecular cloning and characterization of a
 RT brassinosteroid-regulated gene from elongating soybean (Glycine max
 RT L.) epicotyls."
 RL Plant Physiol. 104:161-170(1994).
 CC -1- FUNCTION: POSSIBLE ROLE IN BRASSINOSTEROID-STIMULATED ELONGATION.
 CC -1- SIMILARITY: TO A.THALIANA MER1-5.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L23162; AAA81350.1; -
 DR HSSP: P23904; LAJK.
 DR InterPro: IPR000757; -
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 SQ SEQUENCE 283 AA; 32254 MW; C248810EC7835737 CRC64;

Query Match 8.98; Score 128; DB 1; Length 283;

Best Local Similarity 23.78; Pred. No. 0.0011; Mismatches 73; Indels 34; Gaps 9;

Matches 46; Conservative 41; Indels 34; Gaps 9;
 QY 4 AKDESGAEIYTL-----EEVOYGFPEARKMMA--ASGTYSMFLYONGSEI 48
 DB 45 AKINNGQLSLSDKYSKSGFKSKKEYLFRIDMQLKLVAGNSAGIVTAYTL-----S 98
 QY 49 ADGRPWVEVDIEVLGK--NPGSFOSNITGKAGAKTSEKHAAVSPAADQAFHTYGLEW 105
 DB 99 .SQGPTHEIDIEFEELGNSGDPYILHTNIFTOGKG-NREQPYLWEDPTRN--FHTYSIIW 155
 QY 106 TPNVYRRTVDGQEVKTEGQVSNL--TGQGLRF--NWSSESANVG---QFDESKLP 158
 DB 156 KPOHITLVNDTPTRVFNMAEPLGVPPEPKNOPMRYSSLMNADDMATRGGLVKTWMSKAP 215
 QY 159 LFOFINWVKVKKYT 172
 DB 216 FTAVYRNFKALIEFS 229

Search completed: August 23, 2001, 11:16:31
 Job time: 173 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:13:33 ; Search time 32.87 Seconds
(without alignments)
574.727 Million cell updates/sec

Title: US-09-654-652a-1
Perfect score: 1333
Sequence: 1 MVSADFGSAGELTYLEEVY.....TRKGESFNGQVPRDEPAP 248

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1328	99.6	349	2 A44507	licheninase (EC 3.1.1.4)
2	266.5	20.0	851	2 H84053	endo-beta-1,3-1,4
3	265.5	19.9	252	2 A48378	licheninase (EC 3.1.1.4)
4	247.5	18.6	276	2 I40453	licheninase (EC 3.1.1.4)
5	245.5	18.4	242	1 LXB5	licheninase (EC 3.1.1.4)
6	242.5	18.2	802	2 A36910	licheninase, beta(1,3)
7	238.5	17.9	239	1 A29091	licheninase (EC 3.1.1.4)
8	237	17.8	334	1 S23498	licheninase (EC 3.1.1.4)
9	229.5	17.2	243	1 S15388	licheninase (EC 3.1.1.4)
10	226	17.0	237	1 S11927	licheninase (EC 3.1.1.4)
11	225	16.9	238	1 S10611	licheninase (EC 3.1.1.4)
12	222.5	16.7	242	2 JS0611	endo-1,3(4)-beta-g
13	190.5	14.3	302	2 G84053	hybrid-endo-beta-1
14	182.5	13.7	507	2 S64507	probable membrane
15	180	13.5	268	2 S34804	exon protein - Rhl
16	172	12.9	282	2 T02354	xyloglucan endo-1,
17	171.5	12.9	277	2 H85354	hypothetical prote
18	169.5	12.7	282	2 A85354	hypothetical prote
19	169	12.7	642	2 B72428	laminarinase - The
20	163	12.2	286	2 S71225	xyloglucan endo-1,
21	159	11.9	287	2 T04236	xyloglucan endo-1,
22	154.5	11.6	277	2 S71222	xyloglucan endo-1,
23	150.5	11.3	269	2 S61555	xyloglucan endo-1,
24	147	11.0	467	2 S30839	UHR2 protein - yea
25	146.5	11.0	284	2 T52097	xyloglucan endo-1,
26	141.5	10.6	289	2 T06186	xyloglucan endo-1,
27	140.5	10.5	280	2 T02090	protein T10024.17
28	140.5	10.5	310	2 A86239	probable xylogluca
29	140	10.5	305	2 G84568	

30	139.5	10.5	286	2 T06202	xyloglucan endo-1,
31	138	10.4	422	2 S48564	probable membrane
32	133.5	10.0	286	2 S48201	licheninase (EC 3.1.1.4)
33	132	9.9	295	2 T10210	xyloglucan endo-1,
34	131.5	9.9	292	2 T06201	xyloglucan endo-1,
35	129.5	9.7	292	2 T18265	endo-1,3(4)-beta-g
36	128	9.6	283	2 T07678	xyloglucan endo-1,
37	127.5	9.6	269	2 T05895	xyloglucan endo-1,
38	127	9.5	292	2 T04514	xyloglucan endo-1,
39	125.5	9.4	277	2 T01556	end-xyloglucan tra
40	124	9.3	299	2 D84519	probable endoxyllog
41	123	9.2	296	2 C49539	xyloglucan endo-1,
42	121.5	9.1	292	2 T10211	xyloglucan endo-1,
43	121	9.1	284	2 T06200	xyloglucan endo-1,
44	121	9.1	293	2 T10523	xyloglucan endo-1,
45	120	9.0	877	2 JN0772	glucan endo-1,3-be

ALIGNMENTS

RESULT 1	A44507	1	licheninase (EC 3.2.1.73) - Fibrobacter succinogenes
C:Species:	Fibrobacter succinogenes		
C:Date:	03-Mar-1993	#sequence_revision	03-Mar-1993
C:Accession:	A44507		
R:Teather, R.M.; Erfle, J.D.			
J. Bacteriol. 172, 3837-3841, 1990			
A:Title:	DNA sequence of a Fibrobacter succinogenes mixed-linkage beta-glucanase (1.3		
A:Reference number:	A44507; MUID:90299807		
A:Accession:	A44507		
A:Status:	preliminary		
A:Molecule type:	DNA		
A:Residues:	1-349 <TEA>		
A:Cross-references:	EMBL:M33676; NID:9148575; PIDN:AAA24896.1; PID:9148576		
C:Keywords:	glycosidase; hydrolase; polysaccharide degradation		
Query Match	99.6%; Score 1328; DB 2; Length 349;		
Best Local Similarity	100.0%; Pred. No. 7.9e-105;		
Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	2 VSADFGSAGELTYLEEVYGFKFAKMAAGSVSSMFLYONSGEIDGPRWVVDIEV	61	
DB	25 VSADFGSAGELTYLEEVYGFKFAKMAAGSVSSMFLYONSGEIDGPRWVVDIEV	84	
QY	62 LGKNGSFQSNITIGKAGAKTSEKHNHASPADAFTTGLTWPVNVRTVGOEYRK	121	
DB	85 LGKNGSFQSNITIGKAGAKTSEKHNHASPADAFTTGLTWPVNVRTVGOEYRK	144	
QY	122 TEGGVSNLTGTGIRFLNMSSESANWVGOPDESKLPYFQINWVKYKYPGEGEGSD	181	
DB	145 TEGGVSNLTGTGIRFLNMSSESANWVGOPDESKLPYFQINWVKYKYPGEGEGSD	204	
QY	182 FTLMWTFDFPDGSRMGKGMWTFDGNRVDTLDKNITSRDSMLLALTRKQESFNGVP	241	
DB	205 FTLMWTFDFPDGSRMGKGMWTFDGNRVDTLDKNITSRDSMLLALTRKQESFNGVP	264	
QY	242 RDDEPAP 248		
DB	265 RDDEPAP 271		
RESULT 2	H84053	2	endo-beta-1,3-1,4 glucanase (licheninase) bgls [imported] - Bacillus halodurans (stra
C:Species:	Bacillus halodurans		
C:Date:	01-Dec-2000	#sequence_revision	01-Dec-2000
C:Accession:	H84053		
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H			
Nucleic Acids Res. 28, 4317-4331, 2000			
A:Title:	Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a		

A: Note: sequence extracted from NCBI backbone (NCBI:U31871, NCBI:P31872)
C: Superfamily: endo-1,4-beta-xylanase homology; Thermotoga xylanase A amino-terminal F:4-239/Domain: endo-1,4-beta-xylanase homology <YTL>
F:259-401/Domain: Thermotoga xylanase A amino-terminal repeat homology <YXA>

```
Query Match          18.2%  Score 242.5;  DB: 2;  Length 802;
Best Local Similarity 34.9%  Pred. No. 1.3e-12;
Matches 58; Conservative 21; Mismatches 74; Indels 13; Gaps 6;

QY 7 FSGAEIYLEEVQYGFEPARKMAAASGVSMFYLQNGSEIADGRPVVEVDIEVLGKNP 66
Db 634 YSGGEFRNNYHHGYTCSQMAKNDGVSSFFTYTGPS---DNPDEIDIELGKNT 690
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 67 GSFOSNIITGKAGAKQTSSEKHHAVSPAADQAFHTYGLEMTPNYRWYDVGGEVRKTEGQ 126
Db 691 TVQGFNYNTYTGKHN---EKLYDLGFDSSEALYGFDMQPNYIWMYDGREVYRA---- 743
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 127 VSNLTGOG-LRFNLMSESA-AWYGQDESKLPYFOFINWYKYK 170
Db 744 TQDIPKTPGKIMMANMPEGLIVDDMLKAFN-GRTPHTAHYQWVYTNK 788
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 7
A29091
Licheninase (EC 3.2.1.73) beta - Bacillus amyloliquefaciens
N:Alternate names: 1.3-1,4-beta-D-glucan 4-glucanohydrolase; beta-glucanase;
C:Species: Bacillus amyloliquefaciens
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A29091
R:Hoefmeister, J.; Kurtz, A.; Bortiss, R.; Knowles, J.
Gene 49, 177-187, 1986
A:Title: The beta-glucanase gene from Bacillus amyloliquefaciens shows extensive homology
A:Reference number: A91564; MUID:87192007
A:Accession: A29091
A:Molecule type: DNA
A:Residues: 1-239 <HOF>
A:Cross-references: GB:M15674; NID:g143009; PIDN:AAA87323.1; PID:g143010
A:Experimental source: strain BE20/78
C:Genetics:
A:Gene: bglA
C:Superfamily: Licheninase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match          17.9%  Score 238.5;  DB: 1;  Length 239;
Best Local Similarity 34.3%  Pred. No. 6.2e-13;
Matches 58; Conservative 26; Mismatches 74; Indels 11; Gaps 6;

QY 3 SAKPFGAEIYLEEVQYGFEPARKMAAASGVSMFYLQNGSEIADGRPVVEVDIEVL 62
Db 80 SYNPFDCGENSESVQTYGGLGLEVRMKPAKNGIVSSFFTYGPRF---GTPDEIDIEFL 136
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 63 GKNPFSOSNIITGKAGAKQTSSEKHHAVSPAADQAFHTYGLEMTPNYRWYDVGGEVRKT 122
Db 137 GKDTTKVQFNYYTNGAG---NHEKPADLGPPAANAAYHTYAFDMQPNISIKWYDQ-LKHT 192
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 123 EGGQVSNLTGTQGLRFLNLMSESA-AWYGQDESKLPYFOFINWYKYK 170
Db 193 ATTGPIPAAPGK-IMMANLMNGTGYDDMDGISTYGNV-PIYAHYDMWRKYR 238
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 8
S23498
Licheninase (EC 3.2.1.73) lIcB precursor - Clostridium thermocellum
N:Alternate names: beta-1,3-1,4-glucanase lIcB; lichenase lIcB
C:Species: Clostridium thermocellum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S23498; S22137
R:Schimming, S.; Schwarz, W.H.; Straudenbauer, W.L.
Eur. J. Biochem. 204, 13-19, 1992
A:Title: Structure of the Clostridium thermocellum gene lIcB and the encoded beta-1,3-
cellulases.
```

A:Reference number: S23498; MUID:92155194
 A:Accession: S23498
 A:Molecule type: DNA
 A:Residues: 1-334 <SCCH>
 A:Cross-references: EMBL:X63355; NID:940697; PIDN:CAA44959.1; PID:940688
 C:Genetics:

A:Gene: l1cB
 C:Superfamily: licheninase l1cB; Clostridium cellulase repeat homology
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:1-27/Domain: signal sequence #status predicted <SIG>
 F:28-334/Product: licheninase #status predicted <MAT>
 F:273-296/Domain: Clostridium cellulase repeat homology <CCR1>
 F:308-331/Domain: Clostridium cellulase repeat homology <CCR2>

Query Match 17.8%; Score 237; DB 1; Length 334;
 Best Local Similarity 30.9%; Pred. No. 1.3e-12;
 Matches 80; Conservative 27; Mismatches 94; Indels 58; Gaps 13;

QY 5 KDFSGAELTYLEEVQ-----YKFEARMKMAASGVSSMFLYONGSEIADGRPWVEVD 58
 DB 82 REVGSGVYKSGEXRTKSPFGYGYEVRMAAKNVGIVSSFFTYTGPS---DNNPMDEID 138
 QY 59 IEVLGRKPGSFQSNITIGKAGAKTSEKHNAVSPAADQAFHTYGLFEMTPYVRYMTVDGGE 118
 DB 139 IEFGLKQTTVQFNMVYNGVGG---NEYLHNLGFDASQDFHTYGFEMRPYIDFYVDGKK 195
 QY 119 VRTEGQVSNLCTQG-LRPNLWSSSA-AWVGQFDESKLPLEQFINWVYV----- 169
 DB 196 VYR-----GTNIPYTPCKIMMNLMPGIGVDMLGRYD-GRTPLQAEYKAYYPNGVPOD 250
 QY 170 -----KYTP-----GGGE-GGSDFT-----LDWTDNFTFDGSRMGK 201
 DB 251 NPTPTPIASTPTFNPNPLPKGVNVDGHNSSDYSLEFKRYLLRVIDRFVPGDQS---VA 307
 QY 202 DMTFDGNRVLDTKNITSR 220
 DB 308 DVNRDG-RIDSTDLTMLKR 325

RESULT 9
 S15388
 licheninase (EC 3.2.1.73) - Bacillus licheniformis
 N:Alternate names: endo-beta-1,3-1,4-D-glucanase; lichenase
 C:Species: Bacillus licheniformis
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 R:Liobertas, J.; Perez-Pons, J.A.; Querol, E.
 Eur. J. Biochem. 197, 337-343, 1991
 A:Title: Molecular cloning, expression and nucleotide sequence of the endo-beta-1,3-1,4-
 A:Reference number: S15388; MUID:91224124
 A:Accession: S15388
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-243 <LLO>
 A:Cross-references: EMBL:X57279; NID:939558; PIDN:CAA40547.1; PID:939559
 C:Superfamily: licheninase
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 17.2%; Score 229.5; DB 1; Length 243;
 Best Local Similarity 33.7%; Pred. No. 3.7e-12;
 Matches 57; Conservative 25; Mismatches 76; Indels 11; Gaps 6;
 QY 3 SAKDFGAEIYLEEVQYKFEARMKMAASGVSSMFLYONGSEIADGRPWVEVDIEVL 62
 DB 84 SYNKFDCGERSVQTYGYGLYEVNMPAKNVGIVSSFFTYTGPT---DCTPMDEIDIEFL 140
 QY 63 GKNPGSFQSNITIGKAGAKTSEKHNAVSPAADQAFHTYGLFEMTPYVRYMTVDGGEVRYKT 122
 DB 141 GKTTTVQFNYTNGVGG---NHEKIVNLGFDANSTHTTAFDQPNISIKMYVGG-LKHT 196
 QY 123 EGGOVSNLCTQGLRPNLWSSSA-AWVGQFDESKLPLEQFINWVYV 170

DB 197 ATTQIDPQTPCK--IMMNLWAGVDEWLGSYN-GVTPLSRSLHWRYRK 242

RESULT 10

S11927
 licheninase (EC 3.2.1.73) precursor [validated] - Bacillus macerans
 N:Alternate names: endo-beta-1,3-1,4-glucanase; lichenase
 C:Species: Bacillus macerans
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 15-Sep-2000
 C:Accession: S11927
 R:Boriss, R.; Buettner, K.; Maentzlae, P.
 Mol. Gen. Genet. 222, 278-283, 1990
 A:Title: Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans: homologie
 A:Reference number: S11927; MUID:91109712
 A:Accession: S11927
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-237 <BOR>
 A:Cross-references: EMBL:X59599; NID:9296715; PIDN:CAA39426.1; PID:9296716
 R:Hahn, M.; Heinemann, U.
 Submitted to the Brookhaven Protein Data Bank, December 1994

A:Reference number: A67074; PDB:IMAC

A:Contents: annotation; X-ray crystallography, 2.3 angstroms, residues 26-37, 'P', 39-2
 A:Note: recombinant form with residues 26-41 from Bacillus amyloliquefaciens expresse
 R:Kittel, T.; Simon, O.; Boriss, R.; Heinemann, U.
 Proc. Natl. Acad. Sci. U.S.A. 90, 5287-5291, 1993

A:Title: Molecular and active-site structure of a Bacillus 1,3-1,4-beta-glucanase.
 A:Reference number: A47562; MUID:93281743

A:Contents: annotation; X-ray crystallography, 2.3 angstroms
 A:Note: recombinant form with residues 26-41 from Bacillus amyloliquefaciens expresse

C:Function:

A:Superfamily: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in lichenin an

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:125/Domain: signal sequence #status predicted <PRO>
 F:26-237/Product: licheninase #status experimental <MAT>
 F:55-84/Disulfide bonds: #status experimental

F:128/Active site: Glu #status predicted

Query Match 17.0%; Score 226; DB 1; Length 237;
 Best Local Similarity 35.1%; Pred. No. 7e-12;
 Matches 59; Conservative 23; Mismatches 70; Indels 16; Gaps 8;

QY 7 FSAEALTYLEEVQYKFEARMKMAASGVSSMFLYONGSEIADGRPWVEVDIEVLGKNP 66
 DB 82 FDCAEYRSTNIYGYGLYEVSMKPAKNTGIVSSFFTYTGPT---AHGTQWDEIDIEFLGKDT 138
 QY 67 GSFQSNITIGKAGAKTSEKHNAVSPAADQAFHTYGLFEMTPYVRYMTVDGGEVRYKTEGQ 126
 DB 139 TKYQPNYNTNGVGGH---EVVISLGFDAKSGRFTYAFDQNPQGIKRYVDC-VLKH---A 191
 QY 127 VSNLCTQG-LRPNLWSSSA-AWVGQFDESKLPLEQFINWVYVYKT 172
 DB 192 TANIPSPCKIMMNLWNGTGVDMLSYNGAN-PLVAYEDWV---KYT 235

RESULT 11

S19012
 licheninase (EC 3.2.1.73) precursor - Bacillus polymyxa
 N:Alternate names: endo-beta-1,3-1,4-glucanase; lichenase
 C:Species: Bacillus polymyxa
 C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 18-Jun-1999
 C:Accession: S19012
 R:Gosalbes, M.J.; Perez-Gonzalez, J.A.; Gonzalez, R.; Navarro, A.
 J. Bacteriol. 173, 7705-7710, 1991
 A:Title: Two beta-glucanase genes are clustered in Bacillus polymyxa: molecular cloni
 A:Reference number: S19011; MUID:92041687
 A:Accession: S19012
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-238 <GOS>


```

Db      87  MTLAKRRDNPGLKSNFYIMYKLEVLKLAANGTGVSSFYLDSDLD-----EIDIE 138
QY      61  VLGRNPGSFGQNIITGKAGAKTSEKHHAVPADQAFHTYGLEWTPYVVRWTVGQGEVR 120
Db      139  WVGDNTOFOGNSFTSKGDTTYDGRGEFHGVDPIDK-FHNTLDNMDKTTWYLDGESVR 197
QY      121  KTEGGQVSNLTGTGQ-----LRFNLMS--ESAA---WV-GQFDESKLPLQFTIW 165
Db      198  -----VLSNTSSEGYPOSPMYLMMGIWAGDPDPAAGTIEMAGGETNYNDAPFTMYIEK 251
QY      166  VKVYKYPGQEGSGDFTLWDNFDTFDGSRMGKD 202
Db      252  VIVTDYSTGKKYTGDSGSW-ESIEADGSGIYGRYD 287

```

```

RESULT 15
S34804
exok protein - Rhizobium meliloti megaplasmid 2
C:Species: Rhizobium meliloti
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 26-Aug-1999
C:Accession: S34804; S30072
R:Becker, A.; Kleckmann, A.; Arnold, W.; Puehler, A.
Mol. Gen. Genet. 238: 145-154, 1993
A:Title: Analysis of the Rhizobium meliloti exoX/exoK/exoL fragment: ExoK shows homology
A:Reference number: S34803; MUID:93241147
A:Accession: S34804
A:Molecule type: DNA
A:Residues: 1-268 <BEC>
A:Cross-references: EMBL:Z17219
A:Note: the authors translated the codon GTA for residue 208 as Leu
C:Genetics:
A:Gene: exok
A:Genome: plasmid
C:Superfamily: licheninase

```

```

Query Match      13.5%; Score 180; DB 2; Length 268;
Best Local Similarity 33.8%; Pred. No. 6.3e-08;
Matches 46; Conservative 26; Mismatches 52; Indels 12; Gaps 6;

QY      20  YGKFEARMKMAASGTVSMFLYONGSEIADGRPWVEVDIEVLGKNPQSFQSN-IITGKA 78
      104  YGYEARIKADSGSLNSAFETYIGP--ADKKPHDEIDFEVLGKNYAKVOINQYVSAGK 160
QY      79  GAOKTSEKHHAVPADQAFHTYGLEWTPYVVRWTVGQGEVRK-TEGGQVSNLTGTGLR 137
Db      161  GNEFLAD---VGGAGNGENDYAFWEKNRIRIYVNGELVHEVTDPAKIP--VNAOKIF 214
QY      138  FNLMSSESA-WVGQF 152
Db      215  FSLMGTDLTDMWMTF 230

```

Search completed: August 23, 2001, 11:15:07
 Job time: 94 sec

A:Reference number: S29560
 A:Accession: S29560
 A:Molecule type: mRNA
 A:Residues: 1381 <CAR>
 A:Cross-references: EMBL:X68826; NID:g20716; PIDN:CAA48719.1; PID:g20717
 C:Superfamily: fructose-bisphosphatase
 C:Keywords: phosphoric monoester hydrolase

Query Match 3.2%; Score 8; DB 2; Length 381;
 Best Local Similarity 100.0%; Pred. No. 3.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 SNLTGTGG 135
 |||||

DB 86 SNLTGTGG 93

RESULT 3
 T06408
 Probable fructose-bisphosphatase (EC 3.1.3.11) precursor - garden pea chloroplast
 C:Species: chloroplast Pisum sativum (garden pea)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Jun-1999
 C:Accession: T06408
 R:Hahn, T.R.; Dong, S.M.; Rhim, J.H.
 Submitted to the EMBL Data Library, February 1999
 A:Description: cDNA sequence and red light modulated expression of chloroplast fructose-
 A:Reference number: Z15655
 A:Accession: T06408
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-407 <HAH>
 A:Cross-references: EMBL:L34806; NID:g609560; PIDN:AAD10213.1; PID:g609561
 A:Experimental source: cv. Giant
 C:Genetics:
 A:Genome: chloroplast
 C:Superfamily: fructose-bisphosphatase
 C:Keywords: chloroplast; phosphoric monoester hydrolase
 F:1-50/Domain: transit peptide (chloroplast) #status predicted <TNP>
 F:51-407/Product: fructose-bisphosphatase #status predicted <MNT>

Query Match 3.2%; Score 8; DB 2; Length 407;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 SNLTGTGG 135
 |||||

DB 112 SNLTGTGG 119

RESULT 4
 T07815
 S-locus-specific glycoprotein S8 - radish (fragment)
 C:Species: Raphanus sativus (radish)
 C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
 C:Accession: T07815
 R:Sakamoto, K.; Kusaba, M.; Nishio, T.
 Mol. Gen. Genet. 258, 397-403, 1998
 A:Title: Polymorphism of the S-locus glycoprotein gene (SLG) and the S-locus related gen
 A:Reference number: Z16146; MUID:98311079
 A:Accession: T07815
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-429 <SAK>
 A:Cross-references: EMBL:AB009683; NID:g3327851; PIDN:BAA31730.1; PID:g3327852
 C:Genetics:
 A:Gene: SLG(S8)
 C:Superfamily: S-locus-specific glycoprotein; S-locus-specific glycoprotein homology
 C:Keywords: glycoprotein
 F:32-427/Domain: S-locus-specific glycoprotein homology <SSG>

Query Match 3.2%; Score 8; DB 2; Length 429;
 Best Local Similarity 100.0%; Pred. No. 3.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 213 TDKNIYSR 220
 |||||

DB 253 TDKNIYSR 260

RESULT 5
 D70784
 Probable Rieske [2Fe-2S] iron-sulfur protein cyob - Mycobacterium tuberculosis (strai
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: D70784
 R:Coile, S.T.; Brosch, R.; Parhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
 A:Reference number: A70500; MUID:98295987
 A:Accession: D70784
 A:Molecule type: DNA
 A:Residues: 1-429 <COL>
 A:Cross-references: GB:Z70283; GB:AL123456; NID:g3261561; PIDN:CAA94264.1; PID:e23357
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: qcrA
 C:Superfamily: Rieske [2Fe-2S] homology
 C:Keywords: 2Fe-2S; metalloprotein; Rieske iron-sulfur protein
 F:343-389/Domain: Rieske [2Fe-2S] homology <RSK>
 F:353,355,372,375/Binding site: 2Fe-2S cluster (Cys, His, Cys, His) (covalent) #statu

Query Match 3.2%; Score 8; DB 2; Length 429;
 Best Local Similarity 100.0%; Pred. No. 3.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 RKQGESFN 237
 |||||

DB 335 RKQGESFN 342

RESULT 6
 S55439
 Hypothetical protein ykfF - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 15-Oct-1999
 C:Accession: S55439; E70061
 R:Glaser, P.; Danchin, A.
 Submitted to the EMBL Data Library, May 1995
 A:Description: Cloning and sequencing of the Bacillus subtilis chromosomal region fro
 A:Reference number: S55414
 A:Accession: S55439
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-95 <GIA>
 A:Cross-references: EMBL:Z49782; NID:g853752; PIDN:CAA89886.1; PID:g853778
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C:Bron, S.; Brulliet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A:Enllich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koestler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
 akuchul, M.; Yamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A69580; MUID:98044033
 A:Accession: E70061
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-95 <KUN>
 A:Cross-references: GB:299122; GB:AL009126; NID:92636029; PIDN:CAB15716.1; PID:e1184605;
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: ywxf

Query Match 2.8%; Score 7; DB 2; Length 95;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 RKMMAA 32
 DB 71 RKMMAA 77

RESULT 7
 F84469
 Probable glycine-rich protein [imported] - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: F84469
 R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
 Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-766, 1999
 A>Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
 A:Reference number: A84420; MUID:20083487
 A:Accession: F84469
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-115 <STG>
 A:Cross-references: GB:AE002093; NID:94581171; PIDN:AAD24654.1; GSPDB:GNO0139
 C:Genetics:
 A:Gene: At2g05530
 A:Map position: 2

Query Match 2.8%; Score 7; DB 2; Length 115;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 AAASGV 36
 DB 22 AAASGV 28

RESULT 8
 S64079
 Probable membrane protein YGL072c - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: hypothetical protein G3263
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 09-Jun-2000
 C:Accession: S64079
 R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
 submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64071
 A:Accession: S64079
 A:Molecule type: DNA
 A:Residues: 1-119 <RIB>
 A:Cross-references: EMBL:Z72596; NID:g1322584; PID:e243291; PID:g1322587; GSPDB:GNO0007;
 A:Experimental source: strain 5288C
 C:Genetics:
 A:Gene: MIPS:YGL072C
 A:Map position: 7L
 C:Superfamily: *Saccharomyces cerevisiae* probable membrane protein YGL072c
 C:Keywords: transmembrane protein

F:35-51/Domain: transmembrane #status predicted <TM>

Query Match 2.8%; Score 7; DB 2; Length 119;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 FSGAEY 13
 DB 98 FSGAEY 104

RESULT 9
 F69864
 Hypothetical protein YKTD - *Bacillus subtilis*
 C:Species: *Bacillus subtilis*
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
 C:Accession: F69864
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Gal
 lech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y. M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
 A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
 akuchil, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunori, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
 A:Reference number: A69580; MUID:98044033
 A:Accession: F69864
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-304 <KUN>
 A:Cross-references: GB:299111; GB:AL009126; NID:92633629; PIDN:CAB13342.1; PID:e11850
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yktd

Query Match 2.8%; Score 7; DB 2; Length 304;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 KMAASG 34
 DB 240 KMAASG 246

RESULT 10
 F82672
 APP sulfurylase, small subunit XF1500 [imported] - *Xylella fastidiosa* (strain 9a5c)
 C:Species: *Xylella fastidiosa*
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: F82672
 R:Randomus, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Seq
 Nature 406, 151-157, 2000
 A>Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
 A:Reference number: A82515; MUID:20965717
 A>Note: for a complete list of authors see reference number A59328 below
 A:Accession: F82672
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-317 <SIM>
 A:Cross-references: GB:AE003980; GB:AE003849; NID:g9106531; PIDN:AAF84309.1; GSPDB:GN
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
 as-Neto, E.; Docena, C.; El-Dorfy, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Prohm
J.D.; Jungueira, M.L.; Kemp, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurahoe, E.E.; Laig
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savaas
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1500
C:Superfamily: nodulation protein nodP

Query Match 2.8%; Score 7; DB 2; Length 317;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 220 RDGMIL 226
|||
Db 234 RDGMIL 240

RESULT 11
T08814
1.3-beta-glucanase (EC 3.2.1.-) SGN1 - soybean

C:Species: Glycine max (soybean)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 13-Aug-1999
C:Accession: T08814
R:Cheong, Y.H.; Kim, C.Y.; Chun, H.J.; Hong, J.C.; Bank, J.D.; Cho, M.J.
submitted to the EMBL Data Library, November 1995
A:Description: Isolation and characterization of the soybean beta-1,3-glucanase induced
A:Reference number: Z16477
A:Accession: T08814
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-348 <CHE>
A:Cross-references: EMBL:U01323; NID:G1403674; PID:G1403675
A:Experimental source: strain Williams
C:Genetics:
A:Gene: SGN1
A:Introns: 32/1
A:Note: Induced by wounding and fungal elicitor
C:Superfamily: beta-1,3-glucanase
C:Keywords: glycosidase; hydrolase

Query Match 2.8%; Score 7; DB 2; Length 348;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 TGRGAQ 81
|||||
Db 28 TGRGAQ 34

RESULT 12

S15992
flavohemoglobin hmp - Escherichia coli
N:Alternate names: ferrisiderophore reductase fsrB; flavohemoprotein
C:Species: Escherichia coli
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 03-Mar-2000
C:Accession: S15992; S21161; S25241; G65032
R:Vasudevan, S.G.; Armarego, W.L.F.; Shaw, D.C.; Lilley, P.E.; Dixon, N.E.; Poole, R.K.
Mol. Gen. Genet. 226, 49-58, 1991
A:Title: Isolation and nucleotide sequence of the hmp gene that encodes a haemoglobin-Li
A:Reference number: S15991; MUID:91238719
A:Accession: S15992
A:Molecule type: DNA
A:Residues: 1-396 <YAS>
A:Cross-references: EMBL:X58872; NID:G41730; PID:CAA41682.1; PID:G41731

R:Andrews, S.C.; Shipley, D.; Keen, J.N.; Findlay, J.B.C.; Harrison, P.M.; Guest, J.R
FEBS Lett. 302, 247-252, 1992
A:Title: The haemoglobin-like protein (HMP) of Escherichia coli has ferrisiderophore
A:Reference number: S21161; MUID:92290008

A:Accession: S21161
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-37, 'Q', 39-64:359-396 <AND>
R:Pliemann, M.D.; Stauffer, G.V.
Gene 22, 9-18, 1983

A:Title: Characterization of the Escherichia coli gene for serine hydroxymethyltransf
A:Reference number: S25241; MUID:83235562

A:Accession: S25241
A:Molecule type: DNA
A:Residues: 1-10 <PLA>

A:Cross-references: EMBL:X58872

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617

A:Accession: G65032
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-396 <BLAT>

A:Cross-references: GB:AE000341; GB:U00096; NID:G1788899; PID:AACT5605.1; PID:G17889
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:

A:Gene: hmpA (hmp)

C:Superfamily: Flavohemoglobin; cytochrome-b5 reductase homology; globin homology
C:Keywords: chromoprotein; flavoprotein; heme; iron; metalloprotein; monomer

F:2-136/Domain: globin homology <GLB>

F:157-390/Domain: cytochrome-b5 reductase homology <CBR>

F:53/Binding site: oxygen (Gln) (distal axial ligand) #status predicted

F:85/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 2.8%; Score 7; DB 1; Length 396;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 123 EGGQVSN 129
|||||
Db 225 EGGQVSN 231

RESULT 13

F85900
hypothetical protein hmpA [imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001

C:Accession: F85900

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Iller, L.; Grobeck, E.J.; Davis, N.W.; Llim, A.; Dimantanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: F85900

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-396 <STO>

A:Cross-references: GB:AE005174; NID:G12516962; PID:AA657666.1; GSPDB:GN00145; UNCP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: hmpA

C:Superfamily: Flavohemoglobin; cytochrome-b5 reductase homology; globin homology

Query Match 2.8%; Score 7; DB 2; Length 396;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 123 EGGQVSN 129

|||||||
DB 225 ECGVSN 231

RESULT 14
t36193
probable salicylate hydroxylase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: t36193
R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, March 1999
A:Reference number: 221600
A:Accession: t36193
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-420 <SAU>
A:Cross-references: EMBL:AL051707; PIDN:CAB38889.1; GSPDB:GN00070; SCOEDB:SCE29.14c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCE29.14c
C:Superfamily: tetracycline 6-hydroxylase

Query Match 2.8%; Score 7; DB 2; Length 420;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 226 LALTRKG 232
|||||||
DB 20 LALTRKG 26

RESULT 15
D64186
conserved hypothetical protein H1154 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: D64186
R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Flieschmann, R.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: D64186
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-440 <TIGR>
A:Cross-references: GB:U32795; GB:I42023; NID:g1574708; PIDN:AAC22809.1; PID:g1574711; T
C:Superfamily: Bacillus subtilis sodium-glutamate symporter homolog yhcL

Query Match 2.8%; Score 7; DB 2; Length 440;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 GOVSMLT 131
|||||||
DB 150 GOVSMLT 156

Search completed: August 23, 2001, 11:18:02
Job time: 209 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:15:13 ; Search time 19.47 Seconds
(without alignments)
436.331 Million cell updates/sec

Title: US-09-654-652a-1

Perfect score: 248
Sequence: 1 MVSANDFSGAELTYLEVOY.....TRKGESFNGQVPRDEPAP 248

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size: 0

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	247	99.6	349	1 GUB_FIBSU	P17989 fibroblacter
2	8	3.2	317	1 CYSD_RHIME	P36892 rhizobium m
3	8	3.2	407	1 F16P_PEA	P46275 pisum sativ
4	8	3.2	429	1 OORA_MYCTU	Q10387 mycobacteri
5	7	2.8	95	1 YWKF_BACSU	P45874 bacillus su
6	7	2.8	119	1 YGH2_YEAST	P3161 saccharomyc
7	7	2.8	301	1 NODP_RHITS	P72338 rhizobium s
8	7	2.8	396	1 HMPA_ECOLI	P42332 escherichia
9	7	2.8	440	1 YDJN_HABIN	P45079 haemophilus
10	7	2.8	468	1 ETS2_MOUSE	P15037 mus musculu
11	7	2.8	469	1 ETS2_HUMAN	P15036 homo sapien
12	7	2.8	546	1 BGAM_HUMAN	P16279 homo sapien
13	7	2.8	677	1 BGAL_HUMAN	P16278 homo sapien
14	7	2.8	994	1 MISA_DICDI	P22467 dictyosteli
15	7	2.8	1024	1 SYIP_STAUB	P41368 staphylococ
16	7	2.8	1025	1 SIAP_CAOCR	P35828 caulobacter
17	6	2.4	88	1 WIRA_MHEAT	O01482 triticum ae
18	6	2.4	95	1 WHED_BP1E1	O80294 bacterioph
19	6	2.4	131	1 YIL2_STAUB	P41370 staphylococ
20	6	2.4	132	1 FOLB_MYCLE	O69529 mycobacteri
21	6	2.4	133	1 Y05C_BPT4	P39258 bacterioph
22	6	2.4	137	1 CYB5_ORYSA	P49100 oryza sativ
23	6	2.4	146	1 CYTC_HUMAN	P01034 homo sapien
24	6	2.4	146	1 CYTC_MACMU	O19092 macaca mula
25	6	2.4	148	1 CYTC_BOVIN	P01035 bos taurus
26	6	2.4	157	1 FKBL_METHA	O57726 methanococ
27	6	2.4	158	1 YCBM_BACSU	P42245 bacillus su
28	6	2.4	160	1 VG34_HSVB	P28989 equine herp
29	6	2.4	161	1 YC87_MYCTU	Q10613 mycobacteri
30	6	2.4	174	1 ARF3_DROME	P40946 drosophila
31	6	2.4	174	1 ARF6_CHICK	P26990 gallus galli
32	6	2.4	174	1 ARF6_HUMAN	P26438 homo sapien
33	6	2.4	174	1 ARF6_XENLA	P51645 xenopus lae

34	6	2.4	176	1 VEGP_PIG	P53715 sus scrofa
35	6	2.4	178	1 MYCB_RAT	P15063 rattus norv
36	6	2.4	179	1 UBCH_XENLA	P56616 xenopus lae
37	6	2.4	191	1 TERE_ALCSP	P18782 alcaligenes
38	6	2.4	193	1 FLHC_PROMI	O34202 proteus mir
39	6	2.4	194	1 FLHC_XENNE	O9x9f3 xenorhabdus
40	6	2.4	199	1 PEXH_YEAST	P40155 saccharomyc
41	6	2.4	202	1 Y105_MYCPN	P75528 mycoplasma
42	6	2.4	204	1 PRTB_SCYLI	P15369 scytalidium
43	6	2.4	206	1 YPTL_YEAST	P01123 saccharomyc
44	6	2.4	209	1 RR3_CRATE	P16631 gracilaria
45	6	2.4	215	1 US11_HCMVA	P09727 human cytom

ALIGNMENTS

RESULT	ID	GUB_FIBSU	STANDARD	PRT	349 AA
1	AC	P17989			
DT	01-NOV-1990	(Rel. 16, Created)			
DT	01-NOV-1990	(Rel. 16, Last sequence update)			
DT	15-JUL-1999	(Rel. 38, Last annotation update)			
DE	BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)				
DE	(1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (MIXED LINKAGE BETA-GLUCANASE) (LICHENASE).				
OS	Fibroblacter succinogenes (Bacteroides succinogenes).				
OC	Bacteria; Fibrobacter/Acidobacteria group; Fibrobacter group;				
OX	NCBI_TaxID=833;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 28-57.				
RC	STRAIN-ISOLATE S85;				
RX	MEDLINE-9029807; Pubmed-2193918;				
RA	Teather R.M., Erfle J.D.;				
RT	"DNA sequence of a fibrobacter succinogenes mixed-linkage beta-glucanase (1,3-1,4-beta-D-glucan 4-glucanohydrolase) gene.";				
RL	J. Bacteriol. 172:3837-3841(1990).				
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.				
CC	-1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: M33676; AAA24896.1; -				
DR	PIR: A44507; A44507.				
DR	HSSP; P23904; IAK.				
DR	InterPro: IPR000757; -				
DR	Pfam: PF00722; Glyco_hydro_16; 1.				
DR	PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.				
KW	Hydrolase; Glycosidase; Signal; Repeat.				
FT	SIGNAL	1	27		
FT	CHAIN	28	349		
FT	ACT_SITE	79	79		
FT	ACT_SITE	83	83		
FT	DOMAIN	271	307		
FT	REPEAT	271	277		
FT	REPEAT	278	284		
FT	REPEAT	285	291		
FT	REPEAT	292	298		
FT	REPEAT	301	307		
SO	SEQUENCE	349 AA;	37737 MW;	16DCAF5BDEFCS78A CRC64;	

Query Match 99.6%; Score 247; DB 1; Length 349;
Best Local Similarity 100.0%; Pred. No. 2e-250;

Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VSADFGAELTYLEEVGYGFEARMKAAAGTSSMTLYQNGSEIADGPRWVEVDIEV 61
 |||||
 DB 25 VSADFGAELTYLEEVGYGFEARMKAAAGTSSMTLYQNGSEIADGPRWVEVDIEV 84
 |||||

QY 62 LKKNPSSFGOSNIITGKAGAKTSEKHHAVSPADQAFHTYGLEWTPVYRWTVGQEVK 121
 |||||
 DB 85 LKKNPSSFGOSNIITGKAGAKTSEKHHAVSPADQAFHTYGLEWTPVYRWTVGQEVK 144
 |||||

QY 122 TEGGVSNLTGTOGLRNNLWSSSAAMVGOFPDESKLPFOFINNVKYYKTPGCGEGSD 181
 |||||
 DB 145 TEGGVSNLTGTOGLRNNLWSSSAAMVGOFPDESKLPFOFINNVKYYKTPGCGEGSD 204
 |||||

QY 182 FTLDMTNFTPDGSRKMGKDMFTDGNRVDLTDKNITSRQGLTLALTRKGOSEFNGOVP 241
 |||||
 DB 205 FTLDMTNFTPDGSRKMGKDMFTDGNRVDLTDKNITSRQGLTLALTRKGOSEFNGOVP 264
 |||||

QY 242 RDDEPAP 248
 |||||

DB 265 RDDEPAP 271
 |||||

RESULT 2
 CYS_D_RHIME STANDARD; PRT; 317 AA.
 AC P56892;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE SULFATE ADENYLYLTRANSFERASE SUBUNIT 2 (EC 2.7.7.4) (SULFATE ADENYLYLTRANSFERASE) (SAT) (ATP-SULFURYLASE SMALL SUBUNIT).
 GN Rhizobium meliloti (Sinorhizobium meliloti).
 OS Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxId=382;
 RX MEDLINE=99395034; PubMed=10464198;
 RA Adola A.P., Willis M.G., Wang R.C., Long S.R.;
 RT "Reduction of adenosine-5'-phosphosulfate instead of 3'-phosphosulfate-5'-phosphosulfate in cysteine biosynthesis by Rhizobium meliloti and other members of the family Rhizobiaceae";
 RT J. Bacteriol. 181:5280-5287(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA Galibert F., Capela D., Hubler-Barloy F., Gatius M., Batut J.,
 RA Boistard P., Guzy J., Kahn D., Thebaud P., Goffeau A.,
 RA Purnelle B., Pohl T., Bothe G., Schneider S., Portetelle D.,
 RA Vandenbol M., Puehler A., Becker A., Weidner S.;
 RL Submitted (May-2000) to the SWISS-PROT data bank.
 CC -1- CATALYTIC ACTIVITY: ATP + SULFATE -> PYROPHOSPHATE + ADENYLYLSULFATE.
 CC -1- PATHWAY: FIRST STEP IN THE SULFATE ACTIVATION PATHWAY. THIS REACTION OCCURS EARLY IN THE REDUCTIVE BRANCH OF THE CYSTEINE BIOSYNTHETIC PATHWAY.
 CC -1- SUBUNIT: HETERODIMER COMPOSED OF CYS_D, THE SMALLER SUBUNIT, AND CYSN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY. CYS_D SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: AF158023; AAD55760.1; -
 CC InterPro: IPR002500; -

DR Pfam: PF01507; PAPS_reduct; 1.
 KW Cysteine biosynthesis; Transferase; Nucleotidyltransferase.
 FT CONFLICT 260 260 R -> A (IN REF. 1).
 SQ SEQUENCE 317 AA: 36429 MW: 093F2724D22841CB CRC64;

Query Match 3.2%; Score 8; DB 1; Length 317;
 Best local Similarity 100.0%; Pred. No. 1.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 RDGMLIIA 227
 |||||
 DB 234 RDGMLIIA 241
 |||||

RESULT 3
 F16P_PEA STANDARD; PRT; 407 AA.
 AC P46275; Q37263;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE FRUCTOSE-1,6-BISPHOSPHATASE, CHLOROPLAST PRECURSOR (EC 3.1.3.11)
 DE (D-FRUCTOSE-1,6-BISPHOSPHATE 1-PHOSPHOHYDROLASE) (FBPASE).
 GN FBP.
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 OC Fabales; Fabaceae; Papilionoideae; Pisum.
 OX NCBI_TaxId=3888;
 RX MEDLINE=95175626; PubMed=7870839;
 RA Dong S.M., Rhim J.H., Hahn T.R.;
 RT "Nucleotide sequence analysis of a cDNA encoding chloroplastic fructose-1,6-bisphosphatase from pea (Pisum sativum L.)";
 RL Plant Physiol. 107:313-314(1995).
 RN [2]
 RP SEQUENCE OF 27-407 FROM N.A.
 RC STRAIN=CV. GIANT; TISSUE=leaf;
 RX MEDLINE=94297517; PubMed=7764999;
 RA Carrasco J.L., Cheuca A., Prado F.E., Hermoso R., Lazaro J.J.,
 RA Ramos J.L., Sahravy M., Lopez Gorge J.;
 RT "Cloning, structure and expression of a pea cDNA clone coding for a photosynthetic fructose-1,6-bisphosphatase with some features different from those of the leaf chloroplast enzyme";
 RT Planta 193:494-501(1994).
 RL Planta 193:494-501(1994).
 CC -1- CATALYTIC ACTIVITY: D-FRUCTOSE 1,6-BISPHOSPHATE + H(2)O -> D-FRUCTOSE 6-PHOSPHATE + ORTHOPHOSPHATE.
 CC -1- PATHWAY: THE CHLOROPLAST ISOZYME TAKES PART IN THE REGENERATION OF RIBULOSE BISPHOSPHATE IN THE PHOTOSYNTHETIC CARBON REDUCTION CYCLE (CALVIN CYCLE).
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST STROMA.
 CC -1- INDUCTION: LIGHT ACTIVATION THROUGH PH CHANGES, MG(2+) LEVELS AND ALSO BY LIGHT-MODULATED REDUCTION OF ESSENTIAL DISULFIDE GROUPS VIA THE FERREDOXIN-THIOREDOXIN F SYSTEM (BY SIMILARITY).
 CC -1- MISCELLANEOUS: IN PLANTS THERE ARE TWO FBPAE ISOZYMES: ONE IN THE CYTOSOL AND THE OTHER IN THE CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE FBPAE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: L34806; AAD10213.1; -
 CC EMBL: X68826; CAA48719.1; -
 CC HSSP: P22418; ISPI.

DR InterPro: IPR000146; -
 DR Pfam: PF00316; FBPHASE.1.
 DR PRINTS: PRO0115; FBPHPTASE.
 DR PRINTS: PRO0377; INBPHPTASE.
 DR PROSITE: PS00124; FBPHASE.1.
 KW Hydrolase; Carbohydrate metabolism; Multigene family; Chloroplast;
 FT Transit peptide; Calvin cycle.
 FT TRANSIT 1 50
 FT CHAIN 51 407
 FT ACT_SITE 349 349
 FT DISULFID 223 228
 FT CONFLICT 82 82
 FT CONFLICT 160 160
 FT CONFLICT 247 247
 FT CONFLICT 282 282
 SO SEQUENCE 407 AA; 44511 MW; B60E9164F1F6F9D CRC64;

Query Match 3.2%; Score 8; DB 1; Length 407;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 128 SNLGTGCG 135
 Db 112 SNLGTGCG 119

RESULT 4
 OQRA_MYCTU STANDARD; PRT; 429 AA.
 AC Q10387;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE UBIQUINOL-CYTOCHROME C REDUCTASE IRON-SULFUR SUBUNIT (RIESKE IRON-SULFUR PROTEIN).
 GN OQRA OR RV2195 OR MYCY190.06.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Mycobacteriaceae; Mycobacterium.
 OC Actinomycetales; Corynebacterineae; Mycobacteriidae;
 OC NCB1_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";
 RT Nature 393:537-544(1998).
 RL Nature 393:537-544(1998).
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX. THE RIESKE PROTEIN IS A HIGH POTENTIAL 2FE-2S PROTEIN.
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF THE UBIQUINOL-CYTOCHROME C COMPLEX ARE: CYTOCHROME B, THE RIESKE PROTEIN AND A CYTOCHROME C SUBUNIT.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO RIESKE PROTEINS FROM OTHER SOURCES (MITOCHONDRIA, BACTERIAL, CHLOROPLAST).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: Z70283; CA94264.1; -
 DR Tuberculist; RV2195; -
 DR InterPro: IPR001281; -
 DR PROSITE: PS00199; RIESKE_1; FALSE_NEG.
 DR PROSITE: PS00200; RIESKE_2.1.
 KW Electron transport; Iron-sulfur; Oxidoreductase; Transmembrane.
 FT TRANSMEM 96 116
 FT TRANSMEM 137 157
 FT TRANSMEM 207 227
 FT METAL 353 353
 FT METAL 355 355
 FT METAL 372 372
 FT METAL 375 375
 FT DISULFID 358 374
 SO SEQUENCE 429 AA; 46922 MW; C22C87EC283FE0E8 CRC64;

Query Match 3.2%; Score 8; DB 1; Length 429;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 230 RKQESFN 237
 Db 335 RKQESFN 342

RESULT 5
 YWKF_BACSU STANDARD; PRT; 95 AA.
 AC P45874;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE HYPOTHETICAL 10.1 KDA PROTEIN IN PRA-SPOIIR INTERGENIC REGION.
 GN YWKF.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OC NCB1_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Glaser P., de la Fuente V., Danchin A.;
 RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 75-95 FROM N.A.
 RC STRAIN=168;
 RA Glaser P., Danchin A.;
 RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

QY 26 RMKAAA 32
 Db 71 RMKAAA 77

RESULT 6

YGH2_YEAST STANDARD; PRT; 119 AA.

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE HYPOHETICAL 13.4 KDA PROTEIN IN HSF1-AFT1 INTERGENIC REGION.

YGL072C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OC NCB1_TaxID=4932;

RT Sequence analysis of 203 kilobases from Saccharomyces cerevisiae

RL Yeast 13:1077-1090(1997).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

DR EMBL: Z72596; CA96778.1; -.

DR SGD: S0003040; YGL072C.

KW Hypothetical protein; Transmembrane.

FT TRANSMEM 30 50 POTENTIAL.

SQ SEQUENCE 119 AA; 13388 MW; FID2F040C1C833A6 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 119;

Best Local Similarity 100.0%; Pred. No. 5.9;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 FSGAEY 13

Db 98 FSGAEY 104

RESULT 7

NDOP_RHIS3 STANDARD; PRT; 301 AA.

AC P72338;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE SULFATE ADENYLTRANSFERASE SUBUNIT 2 (EC 2.7.7.4) (SULFATE ADENYLATE

TRANSFERASE) (SAT) (ATP-SULFURYLASE SMALL SUBUNIT) (NODULATION PROTEIN

P).

OS Rhizobium sp. (strain N33).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Rhizobium.

OC NCB1_TaxID=103798;

RN NCB1_TaxID=103798;

RA Cloutier J.;

RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: PROPOSED TO PROVIDE ACTIVATED SULFATE FOR TRANSFER

CC -1- CATALYTIC ACTIVITY: ATP + SULFATE = PYROPHOSPHATE +

CC ADENYLYLSULFATE.

CC -1- SUBUNIT: SULFATE-ACTIVATING ENZYME, NOD AND NODC. MAY BE

CC PHYSICALLY ASSOCIATED (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY. CYSD SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

DR EMBL: U53327; AAB16901.1; -.

DR InterPro: IPR002500; -.

DR Pfam: PF01507; PAPS_reduct; 1.

KW Nodulation; Transferase; Nucleotidyltransferase.

SQ SEQUENCE 301 AA; 35008 MW; 3582F7EAB0861431 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 301;

Best Local Similarity 100.0%; Pred. No. 13;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 RDGMLL 226

Db 218 RDGMLL 224

RESULT 8

HMPA_ECOLI STANDARD; PRT; 396 AA.

AC P24232;

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE FLAVOHEMOGLOBIN (HEMOGLOBIN-LIKE PROTEIN) (FLAVOHEMOGLOBIN)

DE (DIHYDROPTERIDINE REDUCTASE (EC 1.6.99.7)) (FERRISIDEROPHORE REDUCTASE

DE B) (NITRIC OXIDE DIOXYGENASE) (NOD).

GN HMP OR HMPA OR FSRB.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OC NCB1_TaxID=562;

RN NCB1_TaxID=562;

RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.

RC STRAIN-K12;

RC MEDLINE-91238719; PubMed-2034230;

RA Vasudevan S.G., Armarego W.L.F., Shaw D.C., Lilley P.E., Dixon N.E.,

RA Poole R.K.;

RT Isolation and nucleotide sequence of the hmp gene that encodes a

RL haemoglobin-like protein in Escherichia coli K-12.;

RL Mol. Gen. genet. 226:49-58(1991).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-K12;

RC MEDLINE-9749980; PubMed-9205837;

RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,

RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,

RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,

RA Oshima T., Oyama S., Saito N., Saito Y., Sivasubram S.,

RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,

RA Yamagata S., Horiuchi T.;
 RT "Construction of a contiguous 874-kb sequence of the *Escherichia coli*
 RT - K12 genome corresponding to 50,0-68.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113(1997).
 RN [4]
 RP SEQUENCE OF 1-10 FROM N.A.
 RX MEDLINE=83235562; PubMed=6190704;
 RA Plamann M.D., Stauffer G.V.;
 RT "Characterization of the *Escherichia coli* gene for serine
 RT hydroxymethyltransferase.";
 RL Gene 22:9-18(1983).
 RN [5]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RC STRAIN-K12;
 RX MEDLINE=92290008; PubMed=1601132;
 RA Andrews S.C., Shipley D., Keen J.N., Findlay J.B.C., Harrison P.M.,
 RA Guest J.R.;
 RT "The haemoglobin-like protein (HMP) of *Escherichia coli* has
 RT ferriidrophore reductase activity and its C-terminal domain shares
 RT homology with ferredoxin NADP+ reductases.";
 RL FEBS Lett. 302:247-252(1992).
 RN [6]
 RP SEQUENCE OF 1-12.
 RC STRAIN-K12 / EMG2;
 RX MEDLINE=97443975; PubMed=9298646;
 RA Link A.J., Robison K., Church G.M.;
 RT "Comparing the predicted and observed properties of proteins encoded
 RT in the genome of *Escherichia coli* K-12.";
 RL Electrophoresis 18:1259-1313(1997).
 RN [7]
 RP SEQUENCE OF 1-20, AND CHARACTERIZATION.
 RX MEDLINE=98393652; PubMed=9724711;
 RA Gardner P.R., Gardner A.M., Martin L.A., Salzman A.L.;
 RT "Nitric oxide dioxygenase: an enzyme function for flavohemoglobin.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10378-10383(1998).
 CC -1- FUNCTION: POSSESSES DIHYDROPERIDINE REDUCTASE ACTIVITY. ALSO HAS
 CC NITRIC OXIDE DIOXYGENASE ACTIVITY.
 CC -1- FUNCTION: SEEMS TO REDUCE THE COMPLEXED FE+3 IRON OF SIDENOPHORES
 CC TO FE+2, THUS RELEASING IT FROM THE CHELATOR.
 CC -1- CATALYTIC ACTIVITY: NAD(P)H + 6,7-DIHYDROPERIDINE = NAD(P)(+) +
 CC 5,6,7,8-TETRAHYDROPERIDINE.
 CC -1- SUBUNIT: MONOMER.
 CC -1- DOMAIN: CONSISTS OF TWO DISTINCT DOMAINS; ONE IS A HEME-CONTAINING
 CC OXYGEN BINDING DOMAIN IN THE N-TERMINAL REGION AND THE OTHER IS AN
 CC FAD-CONTAINING REDUCTASE DOMAIN FOUND IN THE C-TERMINAL REGION.
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY. TWO DOMAINS
 CC FLAVOHEMOPROTEINS SUBFAMILY.
 CC -1- SIMILARITY: TO A NUMBER OF FAD/NAD(P) FLAVOPROTEIN
 CC OXIDOREDUCTASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; X58872; CAA41682.1; -;
 DR EMBL; AE000341; AAC75605.1; -;
 DR EMBL; D90885; BAA16460.1; -;
 DR EMBL; J01620; AAA23911.1; -;
 DR PIR; S15992; S15992.
 DR PIR; S21161; S21161.
 DR HSP; P04252; IVHB.
 DR Ecogen; EG10456; hmp.
 DR InterPro; IPR000971; -;
 DR InterPro; IPR001433; -;
 DR Pfam; PF00042; globin; 1.
 DR Pfam; PF00175; oxidored_fad; 1.
 DR PROSITE; PS01033; GLOBIN; 1.
 KW Oxidoreductase; NADP; Heme; Flavoprotein; FAD; Iron transport.

FT DOMAIN 1 134 GLOBIN.
 FT METAL 53 53 IRON (HEME DISTAL LIGAND)
 FT METAL 85 85 (BY SIMILARITY).
 FT METAL 85 85 IRON (HEME PROXIMAL LIGAND)
 FT NP_BIND 268 273 MADE (RIBOSE PART) (BY SIMILARITY).
 FT SEQUENCE 396 AA: 43867 MW: 499618DEL1448D68 CRC64;
 SQ
 Query Match 2.8%; Score 7; DB 1; Length 396;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 123 EGGVSN 129
 DB 225 EGGVSN 231
 RESULT 9
 YDUN_HAEIN STANDARD; PRT; 440 AA.
 ID YDUN_HAEIN
 AC P45079;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL SYMPORTER H11154.
 GN H1154.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fletschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McInerney K., Sutton G., Fitchum W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glöck A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fitchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of *Haemophilus*
 RT influenzae Rd.";
 RL Science 269:496-512(1995).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
 CC (SDF). STRONG, TO E.COLI YDUN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; U32795; AAC22809.1; -;
 DR TIGR; H11154; -;
 DR InterPro; IPR001991; -;
 DR Pfam; PF00375; SDF; 1.
 DR PROSITE; PS00713; NA_DICARBOXYL_SYM_1; FALSE_NEG.
 DR PROSITE; PS00714; NA_DICARBOXYL_SYM_2; FALSE_NEG.
 KW Hypothetical protein; Transmembrane; Transport.
 FT TRANSMEM 1 21 POTENTIAL.
 FT TRANSMEM 29 49 POTENTIAL.
 FT TRANSMEM 70 90 POTENTIAL.
 FT TRANSMEM 101 121 POTENTIAL.
 FT TRANSMEM 179 199 POTENTIAL.
 FT TRANSMEM 226 246 POTENTIAL.
 FT TRANSMEM 258 278 POTENTIAL.

FT TRANSMEM 343 363 POTENTIAL.
 FT TRANSMEM 366 386 POTENTIAL.
 FT TRANSMEM 389 409 POTENTIAL.
 SQ SEQUENCE 440 AA; 46491 MW; F9298839559FD06A CRC64;

Query Match 2.88; Score 7; DB 1; Length 440;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 GOVSNTL 131
 DB 150 GOVSNTL 156

RESULT 10
 ETS2_MOUSE STANDARD; PRT; 468 AA.
 AC P15037;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-OCF-2000 (Rel. 40, Last annotation update)
 DE C-ETS-2 PROTEIN.
 GN ETS2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-89042086; PubMed-2847145;
 RA Watson D.K., McWilliams M.J., Lapis P., Lautenberger J.A.,
 RA Schweinfest C.W., Paps T.S.;
 RA "Mammalian ets-1 and ets-2 genes encode highly conserved proteins."
 RL Proc. Natl. Acad. Sci. U.S.A. 85:7862-7866(1988).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: J04103; AAA37581.1; -
 CC PIR: C32066; TVMS2.
 CC HSSP: P14921; 2STW.
 CC TRANSFAC: T01397; -
 CC MGD: MGI:95456; Ets2.
 CC InterPro: IPR000418; -
 CC Pfam: PF00178; Ets.1.
 CC PRINTS: PR00454; ETSDOMAIN.
 CC PROSITE: PS00345; ETS_DOMAIN_1; 1.
 CC PROSITE: PS00346; ETS_DOMAIN_2; 1.
 CC PROSITE: PS00061; ETS_DOMAIN_3; 1.
 CC KMW Proto-oncogene: DNA-binding; Nuclear protein.
 FT DOMAIN 87 170 POINTED.
 FT DNA_BIND 362 442 ETS-DOMAIN.
 SQ SEQUENCE 468 AA; 52827 MW; 5260F3085B7E8B31 CRC64;

Query Match 2.88; Score 7; DB 1; Length 468;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 190 FORFDS 196
 DB 29 FORFDS 35

RESULT 11
 ETS2_HUMAN STANDARD; PRT; 469 AA.
 AC P15036;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-OCF-2000 (Rel. 40, Last annotation update)
 DE C-ETS-2 PROTEIN.
 GN ETS2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-89042086; PubMed-2847145;
 RA Watson D.K., McWilliams M.J., Lapis P., Lautenberger J.A.,
 RA Schweinfest C.W., Paps T.S.;
 RA "Mammalian ets-1 and ets-2 genes encode highly conserved proteins."
 RL Proc. Natl. Acad. Sci. U.S.A. 85:7862-7866(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Zimmermann W.W.K., Korenberg J., Rosenthal A., Schatteroy R.;
 RA Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20289799; PubMed-10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
 RA Onki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
 RA Menzel U., Delabar J., Kumpf K., Lehmann R., Paterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,
 RA Scharte M., Schoen O., Dessario A., Reichelt J., Kauer G., Blocker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Risseisemann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21."
 RL Nature 405:311-319(2000).
 CC
 CC (4)
 CC SEQUENCE OF 399-469 FROM N.A.
 RP MEDLINE-91067187; PubMed-2250910;
 RX Watson D.K., Mavrothalassitis G.J., Jorcyk C.L., Smyth F.E.,
 RA Paps T.S.;
 RA "Molecular organization and differential polyadenylation sites of the
 RT human ETS2 gene."
 RL Oncogene 5:1521-1527(1990).
 RN [5]
 RP SEQUENCE OF 324-469 FROM N.A.
 RX MEDLINE-86042652; PubMed-2997781;
 RA Watson D.K., McWilliams-Smith M.J., Nunn M.F., Duesberg P.H.,
 RA O'Brien S.J., Paps T.S.;
 RA "The ets sequence from the transforming gene of avian
 RT erythroblastosis virus, E26, has unique domains on human chromosomes
 RT 11 and 21: both loci are transcriptionally active."
 RL Proc. Natl. Acad. Sci. U.S.A. 82:7294-7298(1985).
 CC
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 POINTED (PNT) DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: J04102; AAA52412.1; -
 CC DR EMBL: AF017257; AAB94057.1; -
 CC EMBL: AP001732; BAA95514.1; -

RA MEDLINE=91338151; Pubmed=1907800;
RA Yoshida K., Oshima A., Shlimoto M., Fukuhara Y., Sakuraba H.,
RA Yangsiawe N., Suzuki Y.,
RA Human beta-galactosidase gene mutations in GMI-gangliosidosis: a
RT common mutation among Japanese adult/chronic cases.";
RL Am. J. Hum. Genet. 49:435-442(1991).
[17]
RP VARIANT GMI HIS-482.
RX MEDLINE=93138608; Pubmed=1487238;
RA Mosna G., Fattore S., Tubiello G., Brocca S., Trubia M.,
RA Glanzberg E., Gatti R., Danesino C., Minelli A., Piantanida M.,
RT "A homozygous missense arginine to histidine substitution at position
RT 482 of the beta-galactosidase in an Italian infantile GMI-
RT gangliosidosis patient.";
RL Hum. Genet. 90:247-250(1992).
[81]
RN VARIANTS GMI CYS-208; ARG-578; HIS-590 AND GLY-632.
RX MEDLINE=94027054; Pubmed=8218181;
RA Boustany R.-M., Qian W.-H., Suzuki K.,
RT "Mutations in acid beta-galactosidase cause GMI-gangliosidosis in
RT American patients.";
RL Am. J. Hum. Genet. 53:881-888(1993).
[9]
RN VARIANT GMI MET-82.
RX MEDLINE=94256487; Pubmed=8198123;
RA Chakraborty S., Rafi M.A., Wenger D.A.,
RT "Mutations in the lysosomal beta-galactosidase gene that cause the
RT adult form of GMI gangliosidosis.";
RL Am. J. Hum. Genet. 54:1004-1013(1994).
[10]
RN VARIANTS MORQUIO B HIS-83 AND CYS-482.
RX MEDLINE=96049832; Pubmed=7586649;
RA Ishii N., Oohira T., Oshima A., Sakuraba H., Endo F., Matsuda I.,
RA Sukegawa K., Orit T., Suzuki Y.,
RT "Clinical and molecular analysis of a Japanese boy with Morquio B
RT disease.";
RL Clin. Genet. 48:103-108(1995).
[11]
RN VARIANTS GMI HIS-59; ASN-591 AND CYS-591.
RP Mottone A., Bardelli T., Donati M.A., Giorgi M., Di Rocco R.,
RA Gatti R., Taddeucci G., Ricci R., D'Azio A., Zammarchi E.,
RT "Identification of new mutations in six Italian patients affected by a
RT variant form of infantile GMI-gangliosidosis with severe
RT cardiomyopathy.";
RL Am. J. Hum. Genet. 61:A258-A258(1997).
[12]
RN VARIANTS GMI H-59;S-121;C-208;M-240;N-491,& VARIANTS P-10;C-521;G-532
RX MEDLINE=99268417; Pubmed=10338095;
RA Silva C.M.D., Severini M.H., Sopenaa A., Coelho J.C., Zaha A.,
RA d'Azio A., Giugliani R.,
RT "Six novel beta-galactosidase gene mutations in Brazilian patients
RT with GMI-gangliosidosis.";
RL Hum. Mutat. 13:401-409(1999).
[1]
CC -1- FUNCTION: CLEAVES BETA-LINKED TERMINAL GALACTOSYL RESIDUES FROM
CC GALACTOSIDES, GLYCOPROTEINS, AND GLYCOSAMINOGLYCANS.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- SUBCELLULAR LOCATION: LYSOSOMAL.
CC -1- ALTERNATIVE PRODUCTS: BETA-GALACTOSIDASE AND BETA GALACTOSIDASE-
CC RELATED PROTEIN ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME
CC GENE.
CC -1- DISEASE: DEFECTS IN GBI ARE THE CAUSE OF GMI-GANGLIOSIDOSIS. AN
CC AUTOSOMAL RECESSIVE DISORDER WITH THREE MAJOR CLINICAL PHENOTYPES
CC THAT ARE DISTINGUISHED ACCORDING TO THE AGE OF ONSET AND SEVERITY
CC OF SYMPTOMS: INFANTILE, JUVENILE AND ADULT. THE INFANTILE FORM IS
CC RAPIDLY PROGRESSIVE AND RESULTS IN SEVERE CENTRAL NERVOUS SYSTEM
CC DEGENERATION AND VISCEROMEGALY, WITH DEATH USUALLY BETWEEN THE
CC FIRST AND SECOND YEAR.
CC -1- DISEASE: DEFECTS IN GBI ARE THE CAUSE OF MUCCOPOLYSACCHARIDOSIS
CC B (OR MORQUIO B SYNDROME) WHICH IS CHARACTERIZED BY SEVERE BONE
CC DEFORMITIES WITHOUT CNS INVOLVEMENT.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.

CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).-----
DR	EMBL; M27507; AAA51819.1; -;
DR	EMBL; M34423; AAA51823.1; -;
DR	EMBL; M22590; AAA51822.1; -;
DR	PIR; A32611; A32611.
DR	PIR; A31673; A31673.
DR	PIR; A32688; A32688.
DR	PIR; B37086; B37086.
DR	MIM; 230500; -;
DR	MIM; 230600; -;
DR	MIM; 230650; -;
DR	MIM; 253010; -;
DR	InterPro: IPR001944; -;
DR	Pfam: PF01301; Glyco_hydro_35; 1.
DR	PRINTS: PR00742; GLHYDRLASE35.
DR	PROSITE: PS01182; GLYCOSYL_HYDROL_F35; 1.
KW	Hydrolase; Glycosidase; Lysosome; Signal; Alternative splicing;
KW	Glycoprotein; Polymorphism; Disease mutation.
FT	SIGNAL 1 23
FT	PROPEP 24 28
FT	CHAIN 29 677
FT	ACT_SITE 188 188
FT	ACT_SITE 268 268
FT	CARBOHYD 26 26
FT	CARBOHYD 247 247
FT	CARBOHYD 464 464
FT	CARBOHYD 498 498
FT	CARBOHYD 542 542
FT	CARBOHYD 545 545
FT	CARBOHYD 555 555
FT	VARIANT 10 10
FT	VARIANT 49 49
FT	VARIANT 51 51
FT	VARIANT 59 59
FT	VARIANT 82 82
FT	VARIANT 83 83
FT	VARIANT 121 121
FT	VARIANT 123 123
FT	VARIANT 201 201
FT	VARIANT 208 208
FT	VARIANT 240 240
FT	VARIANT 273 273
FT	VARIANT 316 316
FT	BETA-GALACTOSIDASE. PROTON DONOR (POTENTIAL). NUCLEOPHILE (POTENTIAL). N-LINKED (GLCNAC. . .) (POTENTIAL). N-LINKED (GLCNAC. . .) (POTENTIAL). N-LINKED (GLCNAC. . .) (POTENTIAL). N-LINKED (GLCNAC. . .) (POTENTIAL). N-LINKED (GLCNAC. . .) (POTENTIAL). N-LINKED (GLCNAC. . .) (POTENTIAL). N-LINKED (GLCNAC. . .) (POTENTIAL). L -> P. /FTId-VAR_008671. R -> C (IN GM1-GANGLIOSIDOSIS; INFANTILE FORM). /FTId-VAR_003329. I -> T (IN GM1-GANGLIOSIDOSIS; ADULT FORM). /FTId-VAR_003330. R -> H (IN GM1-GANGLIOSIDOSIS; INFANTILE FORM). /FTId-VAR_008672. T -> M (IN GM1-GANGLIOSIDOSIS; ADULT FORM; MILD PHENOTYPE). /FTId-VAR_008673. Y -> H (IN MOROCCO B; 2-5% OF ACTIVITY). /FTId-VAR_008674. R -> S (IN GM1-GANGLIOSIDOSIS; INFANTILE FORM). /FTId-VAR_008675. G -> R (IN GM1-GANGLIOSIDOSIS; INFANTILE FORM). /FTId-VAR_003331. R -> C (IN GM1-GANGLIOSIDOSIS; JUVENILE FORM). /FTId-VAR_003332. R -> C (IN GM1-GANGLIOSIDOSIS; INFANTILE FORM). /FTId-VAR_008676. V -> M (IN GM1-GANGLIOSIDOSIS; INFANTILE FORM). /FTId-VAR_008677. M -> L (IN MOROCCO B; 8% OF ACTIVITY). /FTId-VAR_003333. Y -> C (IN GM1-GANGLIOSIDOSIS; INFANTILE FORM). /FTId-VAR_003334.

Query Match	2.8%	Score 7	DB 1	Length 677
Best Local Similarity	100.0%	Pred. No. 27		
Matches	7	Conservative	0	Mismatches
			0	Indels
			0	Gaps
QY	74	ITKRGGA	80	
Db	465	ITKRGGA	472	

```

RESULT 14
MYSA_DICD1
ID MYSA_DICD1 STANDARD: PRI: 994 AA.
AC P22467;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE MYOSIN IA HEAVY CHAIN (MYOSIN-LIKE PROTEIN ABMA).
GN ABMA OR DMIA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AS-2;
RX MEDLINE=92096486; PubMed=2519618;
RA Titus M.A., Warrick H.M., Spudich J.A.;
RT "Multiple actin-based motor genes in Dictyostelium.";
RL Cell Regul. 1:55-63(1989).
CC -I- FUNCTION: ACTIN-BASED MOTOR PROTEIN, POSSIBLY INVOLVED IN A WIDE
CC RANGE OF MOTILE PROCESSES, SUCH AS CELL MOVEMENT ACROSS A SURFACE,
CC AND EXTENSION AND RETRACTION OF PSEUDOPODIA OR LAMELLIPODIA.
CC -I- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
CC AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -I- SIMILARITY: BELONGS TO THE MYOSIN TYPE I FAMILY (SMALL MYOSINS).
CC -I- SIMILARITY: CONTAINS 2 IQ DOMAINS.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/sib.ch).
CC or send an email to license@sib-sib.ch).
-----
DR EMBL; S73909; AAB20711.1; -.
DR PIR: J00151; J00151.
DR HSSP; P08799; 1AMD.
DR DictyDb; DD01028; abma.
DR InterPro; IPR000048; -.
DR InterPro; IPR001609; -.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR MYOSIN; Actin-binding; ATP-binding; Multigene family.
FT KW DOMAIN; 1 782 GLOBULAR HEAD-LIKE DOMAIN.
FT FT DOMAIN; 783 994 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.
FT FT NP_BIND; 105 112 ATP (POTENTIAL).
FT FT DOMAIN; 575 655 ACTIN-BINDING.
FT FT DOMAIN; 830 931 MEMBRANE-BINDING (POTENTIAL).
SQ SEQUENCE 994 AA; 113286 MW; D30A5D20885B1B8C CRC64;
Query Match 2.8%; Score 7; DB 1; Length 994;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 16 EEOYGYK 22
|||||
Db 694 EEOYGYK 700

```

ID	SYIP_STRAU	STANDARD:	PRT:	1024 AA.
AC	PA1368;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-FEB-1996 (Rel. 33, Last annotation update)			
DE	ISOLEUCYL-tRNA SYNTHETASE, MUPIROCIN RESISTANT (EC 6.1.1.5)			
DE	(ISOLEUCINE-tRNA LIGASE) (ILERS) (MUPIROCIN RESISTANCE PROTEIN).			
GN	MUPR.			
OS	Staphylococcus aureus.			
OC	Plasmid.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Staphylococcus.			
OX	NCBI_TaxId=1280;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN-J2870;			
RC	MEDLINE=94346838; PubMed=8067768;			
RX	Hodgson J.E., Curnock S.P., Dyke K.G.H., Morris R., Sylvester D.R.,			
RA	Gross M.S.;			
RT	"Molecular characterization of the gene encoding high-level mupirocin			
RT	resistance in Staphylococcus aureus J2870.";			
RL	Antimicrob. Agents Chemother. 38:1205-1208(1994).			
RN	[2]			
RP	SEQUENCE OF 504-519 AND 550-609 FROM N.A.			
RP	STRAIN-J2870;			
RC	MEDLINE=91244141; PubMed=1903747;			
RX	Dyke K.G.H., Curnock S.P., Golding M., Noble W.C.;			
RA	"Cloning of the gene conferring resistance to mupirocin in			
RT	Staphylococcus aureus.";			
RT	FEMS Microbiol. Lett. 61:195-198(1991).			
RL	-1- FUNCTION: CONFERS HIGH-LEVEL RESISTANCE TO THE ANTIBIOTIC			
CC	MUPIROCIN (PSEUDOMONIC ACID A), AN ILE-ANALOG THAT COOPTIVELY			
CC	INHIBITS ACTIVATION BY ILE-TRNA SYNTHETASE THUS INHIBITING PROTEIN			
CC	BIOSYNTHESIS.			
CC	-1- CATALYTIC ACTIVITY: ATP + L-ISOLEUCINE + TRNA(ILE) = AMP +			
CC	PYROPHOSPHATE + L-ISOLEUCYL-tRNA(ILE).			
CC	-1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).			
CC	-1- SUBUNIT: MONOMER (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC.			
CC	-1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-tRNA SYNTHETASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL; X75439; CAA53189.1; -;			
DR	EMBL; X59478; CAA42080.1; -;			
DR	EMBL; X59477; CAA42079.1; -;			
DR	InterPro; IPR001412; -;			
DR	InterPro; IPR002300; -;			
DR	InterPro; IPR002301; -;			
DR	Pfam; PF00133; tRNA-synt.1; 1.			
DR	PRINTS; PR00984; TRNASYNTHILE.			
DR	PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.			
KW	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;			
KW	Metal-binding; Zinc; Antibiotic resistance; Plasmid.			
FT	Site 42 52			
FT	Site 585 589			
FT	BINDING 588 588			
FT	CONFLICT 519 519			
FT	CONFLICT 591 593			
FT	SEQUENCE 1024 AA; 118875 MW; 56ECD232CA0C8430 CRC64;			

QY 230 RKQESF 236
| | | | |
Db 29 RKQESF 35

Search completed: August 23, 2001, 11:18:28
Job time: 195 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:13:38 ; Search time 19.66 Seconds

(without alignments)
432.114 Million cell updates/sec

Title: US-09-654-652a-1

Perfect score: 1333
Sequence: 1 MVSADFGSGAEYLLEVOY.....TRKGSEFNGQVPRDEPAP 248

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1328	99.6	349 1	GUB_FIBSU
2	266.5	20.0	259 1	GUB_BACBR
3	245.5	18.4	242 1	GUB_BACSU
4	242.5	18.2	802 1	XYND_RUMFL
5	238.5	17.9	239 1	GUB_BACAM
6	237	17.8	334 1	GUB_CLOTM
7	229.5	17.2	243 1	GUB_BACLI
8	226	17.0	237 1	GUB_PAEMA
9	225	16.9	238 1	GUB_PAEMO
10	197	14.8	269 1	EXOK_RHIME
11	182.5	13.7	507 1	YG46_YEAST
12	147	11.0	347 1	MERS_ARATH
13	147	11.0	347 1	UTR2_YEAST
14	133.5	10.0	286 1	GUB_RHOMR
15	128	9.6	283 1	BRU1_SOYBN
16	117	8.6	397 1	CGKA_ALFCA
17	115	8.6	682 1	EL3B_BACCI
18	105	7.9	465 1	EXSH_RHIME
19	102.5	7.7	879 1	GUNI_CLOTM
20	102.5	7.7	1580 1	ACCR_HUMAN
21	95.5	7.2	1581 1	ACCR_RAT
22	94	7.1	465 1	EGIC_RHIME
23	91.5	6.9	571 1	XYNC_PSEFL
24	90	6.8	490 1	ALGE_PSEAE
25	88.5	6.6	736 1	VP4_ROTPE
26	87.5	6.6	400 1	GUNS_BACAG
27	87	6.5	364 1	VM21_BORHE
28	86.5	6.5	1220 1	CSAC_BACUD
29	86.5	6.5	1385 1	CSAA_BACUD
30	84.5	6.3	409 1	GUN2_BACSD
31	84.5	6.3	1289 1	CSAB_BACUD
32	83.5	6.3	422 1	FBW2_HUMAN
33				Q9ukt8 homo sapien

34	82	6.2	422 1	FBW2_MOUSE
35	82	6.2	1103 1	CHS6_USTMA
36	81.5	6.1	614 1	BRUB_ECOLI
37	81	6.1	411 1	DEOB_LACIC
38	81	6.1	551 1	AMT4_PSESA
39	80.5	6.0	708 1	PAL1_DAUCA
40	79.5	6.0	571 1	PAL4_POPRI
41	79.5	6.0	2628 1	HAGA_PORGI
42	79	5.9	1042 1	EP3_PNECA
43	78.5	5.9	301 1	OMP_ECOLI
44	78.5	5.9	449 1	TBB2_EMENT
45	78.5	5.9	464 1	ACHO_RAT

ALIGNMENTS

RESULT ID	1	GUB_FIBSU	STANDARD:	PRT:	349 AA.
AC	P17989:				
DT	01-NOV-1990 (Rel. 16, Created)				
DT	01-NOV-1990 (Rel. 16, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)				
DE	(1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (MIXED LINKAGE BETA-GLUCANASE) (LICHENASE).				
OS	Fibroblast succinogenes (Bacteroides succinogenes).				
OC	Bacteria; Fibrobacter/Actinobacteria group; Fibrobacter group; Fibrobacter.				
OC	NCB1_TaxID=833;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 28-57.				
RC	STRAIN-ISOLATE S85;				
RX	MEDLINE-90299807; PubMed-2193918;				
RA	Teather R.M., Erle J.D.;				
RT	"DNA sequence of a fibrobacter succinogenes mixed-linkage beta-glucanase (1,3-1,4-beta-D-glucan 4-glucanohydrolase) gene.";				
RL	J. Bacteriol. 172:3837-3841(1990).				
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.				
CC	-1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: M33676; AAA24896.1;				
DR	PIR: A44507; A44507.				
DR	HSSE; P23904; IAJK.				
DR	InterPro: IPR000757;				
DR	Pfam: PF00722; Glyco_hydro.16; 1.				
DR	PROSITE: PS01034; GLYCOSYL_HYDROL.F16; 1.				
KW	Hydrolase; Glycosidase; Signal; Repeat.				
FT	SIGNAL	1	27		
FT	CHAIN	28	349		
FT	ACT_SITE	79	79		
FT	ACT_SITE	83	83		
FT	DOMAIN	271	307		
FT	REPEAT	271	277		
FT	REPEAT	278	284		
FT	REPEAT	285	291		
FT	REPEAT	292	298		
FT	REPEAT	301	307		
SO	SEQUENCE	349 AA;	37737 MM;	16DC4F5BDFC578A CRC64;	

Query Match 99.6%; Score 1328; DB 1; Length 349;
Best Local Similarity 100.0%; Pred. No. 5e-104;

Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VSADFGAELETTLEEVYQGFKEARMKMAASGVSSMFLYONSEIADGRPWVEVIEV 61
 DB 25 VSAKDFSGAELETTLEEVYQGFKEARMKMAASGVSSMFLYONSEIADGRPWVEVIEV 84
 QY 62 LKNGSFQSNITITGKAGAKTSEKHHAVSPADAQFHTYGLEWTPVYVMTVDGQEVK 121
 DB 85 LKNGSFQSNITITGKAGAKTSEKHHAVSPADAQFHTYGLEWTPVYVMTVDGQEVK 144
 QY 122 TEGGVSNLTGTGGLRNNMSSSAAMVQGFDESKLPFOFINMKYKKTTPQGEBSGD 181
 DB 145 TEGGVSNLTGTGGLRNNMSSSAAMVQGFDESKLPFOFINMKYKKTTPQGEBSGD 204
 QY 182 FTLDWTFNPTFGDSRNGKGMFPGDNVLDLTKNYSRQGMILLALTRKGOSEFNGOVP 241
 DB 205 FTLDWTFNPTFGDSRNGKGMFPGDNVLDLTKNYSRQGMILLALTRKGOSEFNGOVP 264
 QY 242 RDDEPAP 248
 DB 265 RDDEPAP 271

RESULT 2

GUB_BACBR ID GUB_BACBR STANDARD; PRT; 259 AA.

AC P37073;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
 GN BGLBB.
 OS Bacillus brevis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Brevibacillus.
 OX NCBI_TaxID=1393;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93159752; PubMed=7763386;
 RA Louw M.E., Reid S.J., Watson T.G.;
 RT "Characterization, cloning and sequencing of a thermostable
 endo-(1,3-1,4) beta-glucanase-encoding gene from an alkalophilic
 RT Bacillus brevis.";
 RT Appl. Microbiol. Biotechnol. 38:507-513(1993).
 CC -1- FUNCTION: HYDROLYSES B-GLUCANS CONTAINING MIXED BETA-1,3 AND
 BETA-1,4 LINKAGES. THIS ENZYME IS THERMOSTABLE; ITS OPTIMAL
 TEMPERATURE IS NEAR 65-70 DEGREES CELSIUS.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
 IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
 CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
 SIMILAR TO LICHENASE OF GERMINATING BARLEY.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: M84339; AAA22265.1; -
 DR HSSP: P23904; IAKK.
 DR InterPro: IPR000757; -
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 DR PROSITE: PRO0737; GLYHYDRLASE16.
 DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
 KW Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 31
 FT CHAIN 32 259 POTENTIAL.
 FT ACT_SITE 142 259 BETA-GLUCANASE.
 FT ACT_SITE 146 146 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 146 146 PROTON DONOR (BY SIMILARITY).

SQ SEQUENCE 259 AA; 29960 MW; A63C09F281FF5D13 CRC64;

Query Match 20.0%; Score 266.5; DB 1; Length 259;
 Best Local Similarity 36.08; Pred. No. 26-15;
 Matches 63; Conservative 24; Mismatches 75; Indels 13; Gaps 6;

QY 3 SAKDFSGAELETTLEEVYQGFKEARMKMAASGVSSMFLYONSEIADGRPWVEVIEV 62
 DB 90 SAKDFSGAELETTLEEVYQGFKEARMKMAASGVSSMFLYONSEIADGRPWVEVIEV 148
 QY 63 GKNPFSQSNITITGKAGAKTSEKHHAVSPADAQFHTYGLEWTPVYVMTVDGQEVK 122
 DB 149 GKNPFSQSNITITGKAGAKTSEKHHAVSPADAQFHTYGLEWTPVYVMTVDGQEVK 205
 QY 123 EGGVSNLTGTGGLRNNMSSSAAMVQGFDESKLPFOFINMKYKKTTPQGEBSGD 175
 DB 206 EGGVSNLTGTGGLRNNMSSSAAMVQGFDESKLPFOFINMKYKKTTPQGEBSGD 253

RESULT 3

GUB_BACSU ID GUB_BACSU STANDARD; PRT; 242 AA.

AC P04957;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
 GN BGLS OR BGL OR LICS OR N15B.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=C120;
 RX MEDLINE=84272222; PubMed=6087283;
 RA Murphy N., McConnell D.J., Cantwell B.A.;
 RT "The DNA sequence of the gene and genetic control sites for the
 RT excreted B. subtilis enzyme beta-glucanase.";
 RT Nucleic Acids Res. 12:5355-5367(1984).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=HL-25;
 RA Tezuka H., Yunki T., Yabuuchi S.;
 RT "Construction of a beta-glucanase hyperproducing Bacillus subtilis
 RT using the cloned beta-glucanase gene and a multi-copy plasmid.";
 RT Agric. Biol. Chem. 53:2335-2339(1989).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSJAL;
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
 RT Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95219081; PubMed=7704256;
 RA Wolf M., Gezel A., Simon O., Boriss R.;
 RT "Genes encoding xylan and beta-glucan hydrolizing enzymes in Bacillus
 RT subtilis: Characterization, mapping and construction of strains
 RT deficient in lichenase, cellulase and xylanase.";
 RT Microbiology 141:281-290(1995).
 RN [5]
 RN SEQUENCE OF 1-6 FROM N.A.
 RC STRAIN=BR151;
 RX MEDLINE=96178961; PubMed=8606172;
 RA Schmetz K., Stuelke J., Gertz S., Krueger S., Krieg M., Hecker M.,
 RA Rak B.;
 RT "LicB, a Bacillus subtilis transcriptional antiterminator protein of
 RT the BglG family.";
 RN J. Bacteriol. 178:1971-1979(1996).
 RN [6]

```

SEQUENCE OF 29-63.
RA YUUKI T., Tezuka H., Yabuchi S.:
RT "Purification and some properties of two enzymes from a beta-glucanase
RT hyperproducing strain, Bacillus subtilis HL-25.";
RL Agric. Biol. Chem. 53:2341-2346(1989).
CC -1- CATALYTIC ACTIVITY: HYDROLAYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
CC -1- MISCELLANEOUS: THE SEQUENCE OF STRAIN 168 IS SHOWN.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
DR EMBL; X00754; CA25328.1; -
DR EMBL; D00518; BAA00405.1; -
DR EMBL; D83026; BAA11697.1; -
DR EMBL; Z46862; CAA86822.1; -
DR EMBL; Z28340; CAA82195.1; -
DR EMBL; Z99124; CAB15943.1; -
DR PIR; A22914; LXBS.
DR PIR; J00110; J00110.
DR HSSP; P27051; 1GBG.
DR Subtilist; B610476; bg1S.
DR InterPro; IPR000757; -.
DR Pfam; PF00722; glyco_hydro_16; 1.
DR PRINTS; PR00737; GLHYDRASE16.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 28
FT CHAIN 29 242
FT MOD_RES 29 29 BETA-GLUCANASE.
FT ACT_SITE 133 133 PYRROLIDONE CARBOXYLIC ACID.
FT ACT_SITE 137 137 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 60 89 PROTON DONOR (BY SIMILARITY).
FT VARIANT 24 24 A -> S (IN STRAIN HL-25).
FT VARIANT 83 83 A -> S (IN STRAIN HL-25).
FT VARIANT 204 204 P -> L (IN STRAIN CJ20).
SQ SEQUENCE 242 AA; 27268 MW; 45958DEA0F22B29 CRC64;

Query Match 18.4%; Score 245.5; DB 1; Length 242;
Best Local Similarity 35.8%; Pred. No. 1,1e-13;
Matches 59; Conservative 24; Mismatches 71; Indels 11; Gaps 6;

OY 7 FSGALYLYLEVOYQKFPARKMAASGVSSMFLYQNGSELDGRPWEVDIEVLGNP 66
Db 87 FDCCGNRSVQRYTGYGLYEVRRKRPANNTGIVSEFFTYPT--DETPEDEIDIEFLGDT 143
OY 67 GSPGSNITITGKAGAKQTESEKIHAAVSPADDAQFHTYGLTEWTPYVYVWYDGOEVRTEGQ 126
Db 144 TKVQENYTYTNGA---NNEKIYDVGFDAAANAYHTTAFDMQPNISINWYDGO LKHTATNQ 199

OY 127 VSNLTGTGLRFLNMSSEA-AWVGQFDESKLPILFQITNWKVYK 170
Db 200 IPTPTGK--IMMNLNMGTVDEMLGSYNGVN-PLVAHYHDWRYTK 241

RESULT 4
XYND_RUMFL STANDARD; PRT; 802 AA.
AC 053317;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE XYLANASE/BETA-GLUCANASE PRECURSOR [INCLUDES: ENDO-1,4-BETA-XYLANASE

```

(EC 3.2.1.8) (XYLANASE); (ENDO-BETA-1,3-1,4 GLUCANASE) (EC 3.2.1.73)
DE [1,3-1,4'-BETA-D-GLOCAN 4'-GLUCANOHYDROLASE] (LICHENASE)].
GN XYND.
OS Ruminococcus flavefaciens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
CC Ruminococcus.
OX NCBI_TaxId=1265;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17;
R8 MEDLINE=93259938; PubMed=8491715;
RA Flint H.J., Martin J., McPherson C.A., Daniel A.S., Zhang J.-X.;
RT "A bifunctional enzyme, with separate xylanase and beta(1,3-1,4)-
glucanase domains, encoded by the xynd gene of Ruminococcus
flavefaciens". 175:2943-2951.(1993).
RL J. Bacteriol. 175:2943-2951.(1993).
CC -1- FUNCTION: CONTAINS TWO CATALYTIC DOMAINS WITH XYLANASE AND ENDO-
BETA-1,3-1,4' GLUCANASE ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: ENDOLYOLYSIS OF 1,4'-BETA-D-XYLIDIC
LINKAGES IN XYLANS.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
IN BETA-D-GLUCANS CONTAINING 1,3'- AND 1,4'-BONDS.
CC -1- PATHWAY: XYLAN DEGRADATION.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
G (FAMILY 11 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 16 OF
GLYCOSYL HYDROLASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).

CC CC
DR EMBL; S61204; AAB26620.1; .
DR HSSP; P23904; IAKK.
DR InterPro; IPRO00157; .
DR InterPro; IPRO001137; .
DR Pfam; PF02018; CBD_6; 1.
DR Pfam; PF00457; Glyco_hydro_11; 1.
DR Pfam; PF00722; Glyco_hydro_16; 1.
DR PRINTS; PR00737; G4HYDLASE16.
DR PRINTS; PR00911; GHYDLASE11.
DR PROSITE; PS00776; GLYCOSYL_HYDROL_F11_1; 1.
DR PROSITE; PS00777; GLYCOSYL_HYDROL_F11_2; 1.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
KW Xylan degradation; Hydrolase; Glycosidase; Signal;
KM Multifunctional enzyme.
FT SIGNAL 1 31
FT CHAIN 32 802 POTENTIAL.
FT DOMAIN 32 244 XYLANASE/BETA-GLUCANASE.
FT DOMAIN 245 523 A (XYLANASE).
FT DOMAIN 524 555 B.
FT DOMAIN 556 802 LINER.
FT ACT_SITE 124 124 C (BETA-GLUCANASE).
FT ACT_SITE 226 226 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 684 684 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 684 684 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 524 529 POLY-THR.
FT DOMAIN 532 543 POLY-THR.
FT DOMAIN 546 553 POLY-THR.
SO SEQUENCE 802 AA; 89091 MM; 2880A689647284AF CRC64;

Query Match 18.2%; Score 242.5; DB 1; Length 802;
Best local Similarity 34.9%; Pred. No. 7.7e+13;
Matches 58; Conservative 21; Mismatches 74; Indels 13; Gaps 6;

7 FSGAEVLTEEFQVKGKFAPRKMAAGSTGVSMPLFYONGSEIADRGPRVEVDIEVIAGNP 66
DB 634 YSGGEFRNNNTYHHGYIGCSMDAMKNDGVASSFFTYTGPS---DDNPDIEDIIELGANT 690

DR Pfam: PF00722; Glyco_hydro.16; 1.
DR PRINTS: PR00737; GLHYDRLASE16.
DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN; 2.
DR PROSITE: PS00448; CLOS_CELLOSOME_RPT; 2.
KM Hydrolyase: Glycosidase; Signal: Repeat.
FT SIGNAL 1 27
FT CHAIN 28 334 POTENTIAL.
FT ACT_SITE 136 136 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 140 140 PROTON DONOR (BY SIMILARITY).
FT DOMAIN 252 269 PRO/THR-RICH (LINKER).
FT DOMAIN 273 331 2 x 24 AA APPROXIMATE REPEATS.
FT REPEAT 273 296 1.
FT REPEAT 308 331 2.
FT CONFLICT 304 334 QSVADVRNRGRIDSTDLTMLKRYLIRAPSL ->
FT P0DCGRHDRVDSGSK (IN REF. 2 AND 3).
SQ SEQUENCE 334 AA; 37897 MW; 0837564E9726F281 CRC64;

Query Match 17.8%; Score 237; DB 1; Length 334;
Best Local Similarity 30.9%; Pred. No. 7.9e-13;
Matches 80; Conservative 27; Mismatches 94; Indels 58; Gaps 13;

QY 5 KDFSGAEIYLEEVQ-----YGFARMKMAASGVYSSMFLYONGSEIADGRPWVEVD 58
DB 82 REYGSGSYKSGEYRTKSPFGYGYEVKMAKNGIVSSFEYTYGFS---DNNFWDEID 138
QY 59 IEVIGKNPFSQSNITITKAGAKQKTSKHNHVSPPADQAFHTYGLEWPNYVRMTVDGOE 118
DB 139 IEFIAKDTTKQVFNWYKMGVCG---NEYLHNLGPDASODEFHTYGEWPRDYIDFVYDGRK 195
QY 119 VRKTEGGGVNLTGTQG-LRFNLMSSESA-AWVGQFDESKLPLPOFIMWVYV----- 169
DB 196 VYR---GTRNIPTVPGKIMNLMVPGIGVDEMLGRYD-GRIPLQAEYEVVYVYRPGVPOD 250
QY 170 -----KTP-----GQGE-GGSDF-----LMDTNEFDTPDGRMGK 201
DB 251 NPTPTPTAPSTPTNPMLPLKGVNDCGHVNSDYSLEKRYLLRVIDREFVQDGS---VA 307
QY 202 DMTFPGNRFVLDLKNYISR 220
DB 308 DVNRDG-RIDSTDLTMLKR 325

RESULT 7
GUB_BACLI STANDARD; PRT: 243 AA.
ID GUB_BACLI
AC P27051;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
GN Bgl.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_Taxid=1402;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9124124; PubMed=2026156;
RA Lloberas J., Perez-Pons J.A., Querol E.;
RT "Molecular cloning, expression and nucleotide sequence of the
endo-beta-1,3-1,4-D-glucanase gene from Bacillus licheniformis.
Predictive structural analyses of the encoded polypeptide.",
RL Eur. J. Biochem. 197;337-343(1991).
RN [2]
RP REVISIONS.
RA Querol E.;
RL Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RX MUTAGENESIS.
MEDLINE=92362869; PubMed=1354172;

RA Planas A., Juncosa M., Lloberas J., Querol E.;
RT "Essential catalytic role of Glu134 in endo-beta-1,3-1,4-D-glucan 4-
glucanohydrolase from B. licheniformis as determined by site-directed
mutagenesis.",
RL FEBS Lett. 308:141-145(1992).
RN [4]
RP MUTAGENESIS.
RX MEDLINE=94237863; PubMed=8182059;
RA Juncosa M., Pons J., Dot T., Querol E., Planas A.;
RT "Identification of active site carboxylic residues in Bacillus
licheniformis 1,3-1,4-beta-D-glucan 4-glucanohydrolase by
site-directed mutagenesis.",
RL J. Biol. Chem. 269:14350-14353(1994).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=96063718; PubMed=7589539;
RA Hahn M., Pons J., Planas A., Querol E., Heinemann U.;
RT "Crystal structure of Bacillus licheniformis 1,3-1,4-beta-D-glucan 4-
glucanohydrolase at 1.8-A resolution.",
RL FEBS Lett. 374:221-224(1995).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4--BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X57279; CAAD0547.1; -
DR PIR; S15368; S15368.
DR PDB; 1GBG; 07-DEC-95.
DR InterPro; IPR000757; -
DR Pfam; PF00722; Glyco_hydro.16; 1.
DR PRINTS: PR00737; GLHYDRLASE16.
DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
KM Hydrolyase: Glycosidase; Signal: 3D-structure.
FT SIGNAL 1 27
FT CHAIN 28 243 POTENTIAL.
FT ACT_SITE 134 134 BETA-GLUCANASE.
FT ACT_SITE 138 138 NUCLEOPHILE.
FT DISULFID 61 90 PROTON DONOR.
FT MUTAGEN 51 51
FT MUTAGEN 89 89 D->N: 30% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 92 92 E->O: 85% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 105 105 E->O: 3% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 105 105 E->O: 50% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 133 133 D->N: 15% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 134 134 E->O: 0.2% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 136 136 D->N: 0.5% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 138 138 E->O: COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 143 143 D->N: 65% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 160 160 E->O: 15% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 168 168 D->N: 60% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 179 179 D->N: 80% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 190 190 D->N: 70% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 219 219 D->N: NO CHANGE IN ACTIVITY.
SQ SEQUENCE 243 AA; 27435 MW; 651188D9AAD609A5 CRC64;

Query Match 17.2%; Score 229 5; DB 1; Length 243;
Best Local Similarity 33.7%; Pred. No. 2.3e-12;
Matches 57; Conservative 25; Mismatches 76; Indels 11; Gaps 6;

QY 3 SAKDFSGAEIYLEEVQYGFARMKMAASGVYSSMFLYONGSEIADGRPWVEVDIEVL 62
DB 84 SYNKRDCGKNSVQYVYGLVFNWYKMGVCG---NEYLHNLGPDASODEFHTYGEWPRDYIDFVYDGRK 140
QY 63 GKNPFSQSNITITKAGAKQKTSKHNHVSPPADQAFHTYGLEWPNYVRMTVDGOEVRKT 122

DB 141 GKDDTKVOFNYYTNGVGV---NHEKIVNLGFDPAANSYHTYAFDQMPNSIKWYDGO-LKHT 196
 OY 123 EGGGVSNLTGNGRLEFNLMSSESA-AMVGOPEDESKLPFOFINWVKYK 170
 DB 197 ATTQIPOTPGK--IMMNLNGAGVDEWLGSYN-GVTPLSRLHWRVYRK 242

RESULT 8
 GUB_PAEPA STANDARD: PRT: 237 AA.

AC P23904;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
 DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
 OS Paenibacillus macerans (Bacillus macerans).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Paenibacillus.
 OX NCBI_TaxID=44252;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91109712; PubMed=2274030;
 RA Borris R., Bueltner K., Maentzsaellae P.;
 RT "Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans:
 RT homologues to other beta-glucanases.";
 RL Mol. Gen. Genet. 222:278-283(1990).
 RN [12]
 RP ACTIVE SITE.
 RX MEDLINE=93094208; PubMed=1360982;
 RA Hoef J.P.B., Condon R., Traeger J.C., Stone B.A.;
 RT "Identification of glutamic acid 105 at the active site of Bacillus
 RT amyloliquefaciens 1,3-1,4-beta-D-glucan 4-glucanohydrolase using
 RT epoxide-based inhibitors.";
 RL J. Biol. Chem. 267:25059-25066(1992).
 RN [13]
 RP MUTAGENESIS OF GLU-128.
 RA Olsen O.;
 RL Thesis (1990), University of Aarhus, Denmark.
 RN [14]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=93281743; PubMed=8099449;
 RA Kettel T., Simon O., Borris R., Heinemann U.;
 RT "Molecular and active-site structure of a Bacillus 1,3-1,4-beta-
 RT glucanase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:5287-5291(1993).
 RN [15]
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RX Hahn M., Kettel T., Heinemann U.;
 RT MEDLINE=96028129; PubMed=7588726;
 RT "Crystal and molecular structure at 0.16-nm resolution of the hybrid
 RT Bacillus endo-1,3-1,4-beta-D-glucan 4-glucanohydrolase H(A16-W).";
 RL Eur. J. Biochem. 232:849-858(1995).
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
 CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
 CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
 CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X55959; CAA39426.1;
 DR PIR: S11927; S11927.
 DR PDB: ZATH; 31-MAR-95.
 DR PDB: IBYH; 31-OCT-93.
 DR PDB: ICPW; 22-JUN-94.

DR PDB: ICPW; 22-JUN-94.
 DR PDB: IMAC; 27-FEB-95.
 DR PDB: IAK; 06-MAY-98.
 DR PDB: IADQ; 06-MAY-98.
 DR InterPro: IPR000757;
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 DR PRINTS: PR00737; GLHYDRLASE16.
 DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
 KW Hydrolase; Glycosidase; Signal; 3D-structure.
 FT SIGNAL 1 23
 FT CHAIN 24 237
 FT ACT_SITE 128 128
 FT ACT_SITE 132 132
 FT DISULFID 55 84
 FT MUTAGEN 128 128
 FT STRAND 26 26
 FT STRAND 29 31
 FT STRAND 41 43
 FT TURN 51 52
 FT STRAND 53 53
 FT STRAND 55 57
 FT STRAND 56 61
 FT HELIX 59 61
 FT STRAND 62 64
 FT TURN 66 67
 FT STRAND 70 78
 FT TURN 79 80
 FT STRAND 81 89
 FT STRAND 93 93
 FT STRAND 96 103
 FT TURN 108 109
 FT STRAND 110 118
 FT HELIX 120 122
 FT TURN 123 123
 FT STRAND 127 134
 FT TURN 135 136
 FT STRAND 137 147
 FT TURN 148 149
 FT STRAND 150 150
 FT STRAND 155 158
 FT TURN 163 164
 FT STRAND 168 174
 FT STRAND 179 183
 FT TURN 184 185
 FT STRAND 186 191
 FT TURN 200 209
 FT TURN 213 216
 FT STRAND 225 236
 SO SEQUENCE 237 AA; 26589 MW; 436EABCDFFC87781 CRC64;

Query Match 17.0%; Score 226; DB 1; Length 237;
 Best Local Similarity 35.1%; Pred. No. 4, 4e-12;
 Matches 59; Conservative 23; Mismatches 70; Indels 16; Gaps 8;

OY 7 FSGAEVLTLEEVYQKFEARKMAAGSVSSMFLYQNGSEIADGPRVVEVIEVIGKRP 66
 DB 82 FDCAEFRSINIYVIGLYEVSMPAKNTGIVSSFTYTG---AHGQNDIEDIEFGKRT 138
 OY 67 GSFQSMITIGKAGAOQRTSEKHNHVAAPADQAFHTYGLWETPNVRYTVGQEVKRTGGO 126
 DB 139 TKVOFNYYTNGVGV---EKVLSIGFDASKGFHTYAFDQMPYIKWYDGO-VLKHT---A 191
 OY 127 VSNLTGTGQ-LRFNLMSSESA-AMVGOPEDESKLPFOFINWVKYK 172
 DB 192 TANIPSTPGKIMMNLNGAGVDEWLGSYNGAN-PLVAEYDWV---KYT 235

RESULT 9
 GUB_PAEPO STANDARD: PRT: 238 AA.
 ID GUB_PAEPO
 AC P45797;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)

```

DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENNO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
GN GLOB.
OS Paenibacillus polymyxa (Bacillus polymyxa).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Paenibacillus.
OX NCBI_TaxID=1406;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 842;
RX MEDLINE=92041687; PubMed=1938968;
RA Gosalves M.J., Perez-Gonzalez J.A., Gonzalez R., Navarro A.;
RT "Two beta-glycanase genes are clustered in Bacillus polymyxa:
RT molecular cloning, expression, and sequence analysis of genes
RT encoding a xylanase and an endo-beta-(1,3)-(1,4)-glucanase.";
RL J. Bacteriol. 173:7705-7710(1991).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES
CC IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.
CC -1- MISCELLANEOUS: BETA-GLUCANASES OF BACILLUS HAVE A SUBSTRATE RANGE
CC SIMILAR TO LICHENASE OF GERMINATING BARLEY.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL; X57094; CAA40379.1; -.
DR HSSP; P23904; ICPM.
DR InterPro; IPR000757; -.
DR Pfam; PF00722; Glyco_hydro_16; 1.
DR PRINTS; PR00737; GLHYDRLASE16.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 26
FT CHAIN 27 238
FT ACET_SITE 129 129 BETA-GLUCANASE.
FT ACET_SITE 133 133 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 56 85 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 238 AA; 26919 MW; C0CFB4EA5D40E8C CRC64;
Query Match 16.9%; Score 225; DB 1; Length 238;
Best Local Similarity 34.5%; Pred. No. 5.4e-12;
Matches 58; Conservative 23; Mismatches 71; Indels 16; Gaps 8;
QY 7 FSGAGELYLEEYOYKGFARMKMAAAGSVSMPLYONGSEIADGRPWVEVDIEYLGNP 66
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 83 FDCGYSRSTNNNGYGLAYEVSMKPRANTGIVSSFYITGSH---GTOMDEIDIEFLGKDT 139
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 67 GSFQSNITTGAKAGAKOTSEKHHAVSPADQAFHTYGLETFPNYVRYMTVDGQEVARTEGQ 126
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 140 TKVQENYNTNGVGH---EKIINLGFDASTSPHTYAFDMQPGYIKWYVDG-VLKHT--A 192
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 127 VSNLTGTQG-LRFNLMSSESA-AWVGQGDSESLPLPQFLINWKKVKKYKT 172
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 193 TTNIPTSTGKIMMNMNGTGYDSWLSGSYNGAN-PLYAEYDWV--KYT 236
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 10
EXOK_RHIME STANDARD; PRT; 269 AA.
AC P33693;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ENNO-1,3-1,4-BETA-GLUCANASE EXOK PRECURSOR (EC 3.2.1.-) (SUCCINOGLYCAN
DE BIOSYNTHESIS PROTEIN EXOK).
GN EXOK.

```

```
0G Rhizobium_melliotti (Slnorhizobium melliotti).
OG Plasmid pSymB (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group.
OX Rhizobiaceae; Slnorhizobium.
RN NCBI_taxid=382;
RP SEQUENCE FROM N.A.
RC STRAIN-1021;
RX MEDLINE-94042869; PubMed-8226645;
RA Glucksmann M.A., Reuber T.L., Walker G.C.;
RT "Family of glycosyl transferases needed for the synthesis of succinoglycan by Rhizobium melliotti." ;
RL J. Bacteriol. 175:7033-7044(1993).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN-RCR2011 / SD47;
RX MEDLINE-93241147; PubMed-8479421;
RA Becker A., Kleckmann A., Arnold W., Puehler A.;
RT "Analysis of the Rhizobium meliloti exoK/exoL fragment: ExoK shows homology to extruded endo-beta-1,3-1,4-glucanases and ExoH resembles membrane proteins." ;
RL Mol. Gen. Genet. 238:145-154(1993).
[3]
RN CHARACTERIZATION.
RP MEDLINE-9826741; PubMed-9560202;
RA York G.M., Walker G.C.;
RT "The Rhizobium meliloti ExoK and Exsh glycanses specifically depolymerize nascent succinoglycan chains." ;
RL Proc. Natl. Acad. Sci. U.S.A. 95:4912-4917(1998).
CC -I FUNCTION: CLEAVES HIGH MOLECULAR WEIGHT SUCCINOGLYCAN TO YIELD LMW SUCCINOGLYCAN. DYNAMICALLY REGULATE THE MOLECULAR WEIGHT DISTRIBUTION OF SUCCINOGLYCAN BY CLEAVING NASCENT SUCCINOGLYCAN ONLY DURING A LIMITED PERIOD AFTER ITS SYNTHESIS, PERHAPS BEFORE IT UNDERGOES A TIME-DEPENDENT CHANGE IN ITS CONFORMATION OR AGGREGATION STATE.
CC -I PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.
CC CC -I SUBCELLULAR LOCATION: SECRETED.
CC CC -I SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/cclib or send an email to license@isb-slb.ch).
-----
DR EMBL: L20758; AAA16048.1; -.
DR EMBL: Z17219; CAAT78927.1; -.
DR HSSP; P23904; IABK.
DR InterPro; IPRO000757; -.
DR Pfam; PF00722; Glyco_hydro_16; 1.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
KW Exopolysaccharide synthetase; polysacchadase; Hydrolase; Signal; Plasmid.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 269 ENDO-1,3-1,4-BETA-GLYCANASE EXOK.
FT COMPLECT 93 102 ACCEGIORPKR -> LRNPDAQG (IN REF. 2) .
SQ SEQUENCE 269 AA; 30083 MW; 073CCEDE5EEF2611 CRC64;
```

DB 200 EVTDPKIP--VNAOKIFFSLMGDTLDMGTF 231

RESULT 11
ID YG46_YEAST STANDARD; PRT; 507 AA.

AC P53301;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOHETICAL 53.8 KDA PROTEIN IN BUB1-HIPI INTERGENIC REGION.
GN YGR189C OR G7553.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NC NCBL_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=97279231; PubMed=9133739;
RA Arroyo J., Garcia-Gonzalez M., Garcia-Saez M.I., Sanchez-Perez M.,
RA Nomdela C.;
RT DNA sequence analysis of a 23,002 bp DNA fragment of the right arm
RT of Saccharomyces cerevisiae chromosome VII.";
RL Yeast 13:357-363(1997).
CC -1- SIMILARITY: SOME, TO YEAST UTR2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 272974; CAA97215.1; -
CC EMBL: X99074; CAA67525.1; -
CC HSSP: P23904; IAOI.
CC SGD: S0003421; CRH1.
CC InterPro: IPR000757; -
CC DR Pfam: PF001722; Glyco_hydro_16; 1.
CC DR Hypothetical protein.
CC FT DOMAIN 63 66 POLY-SER.
CC FT DOMAIN 301 310 POLY-SER.
CC FT DOMAIN 345 357 POLY-SER.
CC FT DOMAIN 387 391 POLY-SER.
CC FT DOMAIN 467 470 POLY-SER.
CC SQ SEQUENCE 507 AA; 52757 MW; 7D7B61F57AE942C CRC64;

Query Match 13.7%; Score 182.5; DB 1; Length 507;
Best Local Similarity 26.7%; Pred. No. 4.7e-08;
Matches 58; Conservative 37; Mismatches 91; Indels 31; Gaps 8;

QY 1 MWSAKDFSGAELTYLEVOYGFKFEARKMAAAGTSSMFLYONGSEIADGRPWVEVDIE 60
DB 87 MTLAKRYDNPSPKSNFYIMGKLEVLKANGIYISFLLQDDDD-----EIDIE 138
QY 61 VLCKNGSFQSNITTKAGAKQKSEKHAAVSPADQAFTTYGLEWTPNRYRWTVGQOEVR 120
DB 139 WVGGDWTFQFSNFSGDTTYYDRGEFHGVDPPTDK-FHNYYTLDMAMDKTWWYLDSESVR 197
QY 121 KTEGGOVSNLTGQGLF-----LRFNLMS-----ESAA-----WV-GQFDESKLLPQFINW 165
DB 198 -----VLSTSTSEGYPOSPMYLMGIMAGDPDNNAGTIEMAGETNVDAPFTWIEK 251
QY 166 VKVYKYYTPGGESESDFTLDMTDFPDGSRWKGCD 202
DB 252 VIATDYSTGKKYTYGQSGSW-ESIEADGSIYGRYD 287

RESULT 12

MERS_ARATH
ID MERS_ARATH STANDARD; PRT; 247 AA.

AC P24806;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE MERRI-5 PROTEIN.
GN MERRI-5.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
NC NCBL_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93005704; PubMed=1840916;
RA Medford J.I., Elmer J.S., Klee H.J.;
RT "Molecular cloning and characterization of genes expressed in shoot
RT apical meristems.";
RL Plant Cell 3:359-370(1991).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN SHOOT APICAL MERISTEMS, ALSO
CC FOUND IN SEEDLINGS AND MERISTEMS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M63166; AAA32828.1; -
CC PIR: J01022; J01022.
CC HSSP: P23904; IAOI.
CC InterPro: IPR000757; -
CC DR Pfam: PF001722; Glyco_hydro_16; 1.
CC SQ SEQUENCE 247 AA; 28295 MW; 831E8441564B45E8 CRC64;

Query Match 11.0%; Score 147; DB 1; Length 247;
Best Local Similarity 28.0%; Pred. No. 1.9e-05;
Matches 44; Conservative 29; Mismatches 60; Indels 24; Gaps 7;

QY 8 SGAEELTYLEVOYGFKFEARKMAA--ASGTSSMFLYONGSEIADGRPWVEVDIEVIGK- 64
DB 54 SSGSFQSKTYLEFGKIDMQIKLVPAGNSAGVTTFYKSEGS-----TWDEIDFEFLGNA 107
QY 65 --NPGSFQSNITTKAGAKQKSEKHAAVSPADQAFTTYGLEWTPNRYRWTVGQ--EVR 120
DB 108 SGDPYTLHTWVYTGKG-DKEQGFHLMFDEPTAN--PFTYSILNPPRIILTVDTHPLESL 164
QY 121 KTEGGOVSNLTGQGLF-----NLMSSESAANV 149
DB 165 KTWLSLVSCFQRTKPRMVRQFMERQSDWATKRRL 201

RESULT 13

ID UTR2_YEAST STANDARD; PRT; 347 AA.

AC P32623;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE UTR2 OR YEL040W OR STGP-ORF18.
GN UTR2 OR YEL040W OR STGP-ORF18.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NC NCBL_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Mulligan J.T., Dietrich F.S., Hennessey K.M., Sehl P., Komp C.,
RA Wei Y., Taylor P., Nakahara K., Roberts D., Davis R.W.;

DT 01-OCT-1996 (Rel. 34, last annotation update)
DE BRASSINOSTEROID-REGULATED PROTEIN BRU1.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 1;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Epicotyl;
RX MEDLINE=94159788; PubMed=8115544;
RA Zurek D.M., Clouse S.D.;
RT Molecular cloning and characterization of a
RT brassinosteroid-regulated gene from elongating soybean (Glycine max
RT L.) epicotyls.*
RL Plant Physiol. 104:161-170(1994)
CC -!- FUNCTION: POSSIBLE ROLE IN BRASSINOSTEROID-STIMULATED ELONGATION.
CC -!- SIMILARITY: TO A.THALIANA MERI-5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L22162; AAA81350.1; -.
DR HSSP: P23904; 1AJK
DR InterPro: IPR000757; -
DR Pfam: PF00722; Glyco_hydro_16; 1.
SQ SEQUENCE 283 AA; 32254 MW; C248810EC7835737 CRC64;

Query Match 9.6%; Score 128; DB 1; Length 283;
Best Local Similarity 23.7%; Pred. No. 0.00085;

Matches 46; Conservative 41; Mismatches 73; Indels 34; Gaps 9;

```
QY 4 AKDFSGAELTYL-----EEVOYGFKFEARKMAA--ASGVSSMFLYONGSEI 48
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 45 AKTFNGQQLSLSDKXVSGSGFKSKKEYLFGRIDMQLKIYAGNSAGTAYTL-----S 98
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 49 ADGRPWVEVDIEVLGK---NPGSFQSNITGKAGAOKTSEKHAAVSPADQAEHTYGLEW 105
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 99 SOGPTHEIDIEFELGNSGDPYILHTNIFTQKG-NREQOFLWFDPTRN--FHTYSIIW 155
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 106 TPNVVRKTVGQGVKRTGEGQVSNL--TGQGLRF--NLWSSGSAWVG---QEDSKLP 158
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 156 KPQHIIPLVNTPIRVFKNAEPLGVPPKNQPMKRIYSSLMNADWDATRGGLVKTWSKAP 215
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 159 LFOFINWKKVKKYT 172
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 216 FTAVYRNFKALIEFS 229
```

Search completed: August 23, 2001, 11:16:30
Job time: 172 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 13:37:21 ; Search time 2905.14 Seconds
(without alignments)
4280.713 Million cell updates/sec

Title: US-09-654-652a-5

Perfect score: 804

Sequence: 1 atggttagcgcgaagaattt.....accacacacacacacactga 804

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_vtg1:*
61: gb_vtg2:*
62: gb_vtg3:*
63: gb_vtg4:*
64: gb_vtg5:*
65: gb_vtg6:*
66: gb_vtg7:*
67: gb_vtg8:*
68: gb_vtg9:*
69: gb_vtg10:*
70: gb_vtg11:*
71: gb_vtg12:*
72: gb_vtg13:*
73: gb_vtg14:*
74: gb_vtg15:*
75: gb_vtg16:*
76: gb_vtg17:*
77: gb_vtg18:*
78: gb_vtg19:*
79: gb_vtg20:*
80: gb_vtg21:*
81: gb_vtg22:*
82: gb_vtg23:*
83: gb_vtg24:*
84: gb_vtg25:*
85: gb_vtg26:*
86: gb_vtg27:*
87: gb_vtg28:*
88: gb_vtg29:*
89: gb_vtg30:*
90: gb_vtg31:*
91: gb_vtg32:*
92: gb_vtg33:*
93: gb_vtg34:*
94: gb_vtg35:*
95: gb_vtg36:*
96: gb_vtg37:*
97: gb_vtg38:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	742.8	92.4	1426	3	FTBGUC
2	64.8	8.1	4411	9	AR069366
3	64.8	8.1	5443	9	AR069362
4	64.8	8.1	5616	9	AR069364
5	61.4	7.6	2735	9	AR054309
6	55	6.8	6363	9	AR069367
7	50	6.2	1416	2	BSPBGA1G
8	49.6	6.2	1234	56	SYHA78M

9	48.6	6.0	98	9	AX026763	AX026763 Sequence
10	47.4	5.9	927	9	AX000165	AX000165 Sequence
11	45.8	5.7	306	56	ASTCEL90	Y09275 Artificial
12	44.6	5.5	1228	56	SYNA0M	Z25873 Synthetic B
13	44.6	5.5	1234	56	SYNA12Y13	Z25876 Synthetic B
14	44.6	5.5	1224	56	SYNA16M	Z25878 Synthetic B
15	44.6	5.5	1234	56	SYNA2M	Z25879 Synthetic B
16	44.6	5.5	1234	56	SYNA36M	Z25880 Synthetic B
17	44.6	5.5	1234	56	SYNA4M	Z25881 Synthetic B
18	44.6	5.5	1234	56	SYNA8M	Z25882 Synthetic B
19	44.6	5.5	1237	56	SYNA12MX	Z25875 Synthetic B
20	43.4	5.4	117	56	ASTOOL10	Y09270 Artificial
21	43.2	5.4	1541	2	BACBGLB	M84339 B. brevis en
22	43	5.3	408	56	ASAVTOVAC	Y09276 Artificial
23	43	5.3	852	2	BMBGLUC	X55959 B. macerans
24	43	5.3	5549	56	AF097413	AF097413 Expressio
25	42.6	5.3	586	9	AX025776	AX025776 Sequence
26	42.2	5.2	1234	56	SYNA107M	Z25874 Synthetic B
27	42.2	5.2	1240	9	A00896	A00896 Bacillus sp
28	42.2	5.2	1240	10	I15724	I15724 Sequence 1
29	42	5.2	357	9	AX025780	AX025780 Sequence
30	42	5.2	6619	9	AR028564	AR028564 Sequence
31	40.8	5.1	1607	2	BLBS1	X57279 B. lichenif
32	39.4	4.9	906	13	AF03327	AJ223327 Aspergill
33	39.4	4.9	1546	12	AF062651	AF062651 Aspergill
34	39	4.9	2492	3	S61204	S61204 xynD-xylana
35	39	4.9	11181	3	REFL32472	AJ132472 Ruminococ
36	38.8	4.8	1815	9	A79318	A79318 Sequence 6
37	38.8	4.8	1815	9	AR074521	AR074521 Sequence
38	38.8	4.8	5469	56	AF117644	AF117644 Expressio
39	38.8	4.8	5631	56	AF117645	AF117645 Expressio
40	38.8	4.8	5727	56	AF116269	AF116269 Expressio
41	38.6	4.8	1204	7	PICINVOIA	M3441 P1g involuc
42	38.4	4.8	2539	6	DMTID56	X77822 D. melanog
43	38.4	4.8	2812	6	DMROT57	X77821 D. melanog
44	38.4	4.8	7047	6	DMTINO	X95241 D. melanog
45	38.4	4.8	10289	6	DMCH2RA	X98094 D. melanog

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES
1	FIBGLUC	F. succinogenes 1,3-1,4-beta-D-glucan 4-glucanohydrolase gene, complete cds.	M33676	M33311	M33676.1	GI:148575	F. succinogenes (strain S85) DNA, clone PJ15.	1	(bases 1 to 1426)	Teather, R.M. and Erile, J.D.	DNA sequence of a fibroblast succinogenes mixed-linkage beta-glucanase (1,3-1,4-beta-D-glucan 4-glucanohydrolase) gene	Draft entry and computer-readable sequence for (1) kindly submitted by R.M. Teather, 11-APR-1990.	Location/Qualifiers
													1. .1426
													/organism="Fibroblast succinogenes"
													/db_xref="taxon:833"
													62. .66
													85. .90
													132. .137
													/note="ribosome binding site"
													145. .1194
													/note="1,3-1,4-beta-D-glucan 4-glucanohydrolase precursor (EC 3.2.1.73)"

QY	4	glttagcgcaaggatlttagcgtgctccgaactctacacgtttagaagatcagtcagtc	92.4% Score 742.8; DB 3; Length 1426; Best Local Similarity 99.7%; Pred. No. 2.4e-189; Matches 744; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
DB	217	gTTAGGCCAAGGATTTTACGGTCCGACGCTACAGCTTAGAGAGATTCAGTACG	
QY	64	aagtttgaaagccgtatgaaagatggcagccgcatcggaacagtcagttcattcttc	
DB	277	AAGTTTAAACCCGTATGAAGATGGCAGCGCATCGGGAACATTCATTCATGTTCTC	
QY	124	taccagaatgtgttcggaatcgcgcgcatgtagaagccctgggttagaagtgatgagtt	
DB	337	TACCAAGATGTTCCGAATTCGCGATGCAAGGCCCTGGGTAGAGTGATTTAGATT	
QY	184	ctcggaagaatccggcgagtttccagttccaaatcattacccgtaagcgcgacacaa	
DB	397	CTCGGCAAGATCCGGGCGAGTTTCCAGTCCACATCATTCACGTTAGCGCGGACAA	
QY	244	aagactagcaaaagacacatgctgttagcccgccgcaatgaagtttccacactac	
DB	457	AAGACTAGCAAAAGCACCATGCTTTAGCCCGCCCATCATGAGCTTTCACACCTAC	
QY	304	ggttcgaatggaactcgaatcgttcgctgtagctgttagcgttcggaagtcgcgaag	
DB	517	GGTTCGAATGCACTCGAATTTACGTCCTGGACCTTACGTCAGAGTCCGCAAG	
QY	364	acggaagtgccaggtttccaaacttgacaggtacacaggaagtcggtttaaacttgg	
DB	577	ACGGAAGTGGCCAGGTTTCCAACTTGACAGGTACACAGGACCTCCGTTTAACTTGG	
QY	424	tcgttcgaagagtcggttgggttggtgacattcgatgaatcaagcttccgctttccag	
DB	637	TGCTTCGAGAGTCGGCTTGGGTTGGCCACTTCGATGAACAAAGCTTCGGCTTTTCAG	
QY	484	ttatcaactgggttcaaggttttaagatacagcgcggaagcggaagcggaagcgac	
DB	697	TTCATCACTGAGGTCAGGTTTAAAGTATACCGCGGCAAGGCAAGGCGGACGAGC	
QY	544	tttagcgttactggaacgaatlttgacaggttgatggtctccgctggtggcaaggt	
DB	757	TTTACGCTTACTGACCGACCAATTTTGAACAGTTTGAATGCTCCGCTGGGCAAGGCT	
QY	604	gacttgacatttgaagtgtaacggtgttcgaactcaacgaagaacatctctccaagat	
DB	817	GACTTGACATTTGACGATTAACGCTGTCGACCTCACGAAGAACATCTTCAAGAGAT	
QY	664	ggcatgttagctcgcgcctccacccgcaaggtcaggaagcttcaagcgcaaggttccg	
DB	877	GGCATGTGATCTCCGCCCTCACCCGCAAGGTCAGGAAGCTTCAACGGCCAGTTCCG	
QY	724	agagatgacgaactgtctccgaattc 749	


```

/translation="MKKKLVLFSCMLLFFSLGISGLVQSPQVAEAEARPIGTAIVETE
ESYDSEKRSKAGVWSNCGMFNATWYPEQVTFESNGKMKLQIEKEDNETASPPYKAGEL

```

TNDYFIYHGLEFVSNMRAKSGTGVSEFFITTGPMDMENMPWDEIDIEFGKDTKIQEN
YFTNGVGNNEHYHELGFDADDFNYAFPMRPESIRWFENGEVHTHTENIPQPKIT
MNNLWPGIGVDGWTGRNGEDTPVVTQIDWVKYTPLEELGCTINEKNKNKYKKCKTKYK

JOURNAL
TITLE
AUTHORS
REFERENCE
1 (bases 1 to 125)
Olsen, O., Borris, R., Simon, O. and Thomsen, K.K.
Hydrid Bacillus (1-3,1-4)-beta-glucanases: engineering thermostable
enzymes by construction of hybrid genes
Mol. Gen. Genet. 225 (2), 177-185 (1991)

JOURNML	Mol. Gen. Genet. 225 (2), 177-185 (1991)
MEDLINE	91172113
REMARK	(sites)
REFERENCE	4 (bases 1 to 1234)

AUTHORS	Pollitz,O., Simon,O., Olsen,O. and Borris,R.
TITLE	Determinants for the enhanced thermostability of hybrid
JOURNAL	(1-3,1-4)-beta-glucanases
MEDLINE	Eur. J. Biochem. 216 (3), 829-834 (1993)
REFERENCE	94009045
AUTHORS	5 (bases 1 to 1234)
TITLE	Pollitz,O.
JOURNAL	Direct Submission
FEATURES	Submitted (26-AUG-1993) Pollitz O., Humboldt University of Berlin, Biology, Warschauer Str. 43, BERLIN, Germany, D-10243 location/Qualifiers
source	1..1234
	/organism="synthetic construct"
	/db_xref="taxon:32630"
source	/focus
	1..1234
	/organism="Bacillus amyloliquefaciens"
	/db_xref="taxon:1390"
source	1..777
	/organism="Bacillus amyloliquefaciens"
	/db_xref="taxon:1390"
source	1..1234
	/organism="Paenibacillus macerans"
	/db_xref="taxon:44252"
sig_peptide	469..543
CDS	469..1188
	/function="endo-1,3-1,4-beta-glucanase"
	/note="Hybrid between bgla (Pos. 1 to 777) and bglm (pos. in Hybrid 778 to 1234, Pos. original 396 to 852)"
	/codon_start=1
	/transl_table=1
	/label=HA78M
	/evidence="experimental"
	/product="hybrid-endo-1,3-1,4-beta-glucanase"
	/protein_id="CA81101.1"
	/db_xref="GI:398203"
	/translation="MKRVLLILVTGLFMSLCGITSVSQAOTGSFEFPENSYNSGLMO KAGYSDGNMFNCIRANNVSNMTSLGEMRIALSPSYNKFDCGNRSVQYVGLEYEV RMPKAKTGVIVSFETTYGPAHGOMEDIIEFLGKDTKVQFNYYNGVGHKVKVIS LGFPSKSGFTVAFDMDPGYIKWYVDVLKHTATANIPISTPGKIMNLMMGTGDVDM LSVNGANPLIAEYDWMVKTSN"
source	778..1234
	/organism="Paenibacillus macerans"
	/db_xref="taxon:44252"
BASE COUNT	373 a 224 c 281 g 356 t
ORIGIN	
Query Match	6.2%; Score 49.6; DB 56; Length 1234;
Best Local Similarity	51.8%; Pred. No. 0.011;
Matches 174; Conservative	0; Mismatches 144; Indels 18; Gaps 2;
OY	13 aagattttagcgtgccaactctaacgcgtttagaagaagtctgaatgaatttaa 72
DB	712 AACAACTTTGAACGGGGGAAAAACCGCTCGGCTCAACAATATGGCTAAGCATTTATATA 771
OY	73 gcccgatatgaagatbgccgcgcgatccgggaacagtcagctcatgctctaccagaat 132
DB	772 GTCCAAATGAAGGCCCAAAAATAATCAGAGATTGTCTCATCTTTTCACGTA----- 824
OY	133 ggttcgaataatgcgcgatggaagcgcttggttagaagtgsatatatgaattctcgcgaag 192
DB	825 --TACAGAGACTGCCTCATGCGCACACAAATAGGGATGAATAATGAAATTCTTAGAANA 882
OY	193 aatccggcagtttcaatccaatacatattaccggttaaggccgcggcacaaaagaactgc 252
DB	883 GACACGACAAAAGTCCAATTTAATATTATACCAATTGGGGTTGGCGGTAT----- 933
OY	253 gaaagacaacatgctgtttaagccccgcgcgcgaatcagagcttccacaactacgctcgaa 312
DB	934 GAAGAAGTTATCTCTTGCGCTTTGATGATGATCAACAAAGGCTTCATACCTATGCTTGCAAT 993
OY	313 tggactccgaattacgctcgcgtggaactgltgaacggt 348

```

Db      994   TGGCAGCCAGGCGTATTAATTCGAATCTACACGCT    1029
RESULT          9
AX026763                               PAT                16-SEP-2000
LOCUS                                     AX026763         98 bp     DNA
DEFINITION Sequence 7 from Patent WO0040712.
ACCESSION AX026763
VERSION   AX026763.1 GI:10187902
KEYWORDS
SOURCE      .
ORGANISM    synthetic construct.
            artificial sequence.
REFERENCE   1 (bases 1 to 98)
AUTHORS    Tavfik,D. and Griffiths,A.
TITLE       Optical sorting method
MEDICAL RES COUNCIL (GB) : TWAKFIK DAN (GB) ; GRIFFITHS ANDREW (GB)
FEATURES
source
1..98
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="per-23d FLAG polylinker"
3...29
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAC09112.1"
/db_xref="GI:10187903"
/translation="MDYKDDDDK"
26 c 26 g 16 t

BASE COUNT      30 a 26 c 26 g 16 t
ORIGIN

Query Match           6.0%: Score 48.6; DB 9; Length 98;
Best Local Similarity 92.7%; Pred. No. 0.022;
Matches 51; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy      731   acgaacctgcctccgagatcgagctcgtcgcgacaaacttgaggccgactcgaga 785
        ||||| |
Db      44   AGGTACCGGATTCGCAATTGCAGACTCCGTCGACAGAAGCTTGCGGCCGCACTCGAGCA 98

RESULT          10
AX000165                               PAT                10-MAR-2000
LOCUS                                     AX000165         927 bp     DNA
DEFINITION Sequence 2 from Patent W09906573.
ACCESSION AX000165
VERSION   AX000165.1 GI:7240678
KEYWORDS
SOURCE      unidentified.
            unclassified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 927)
AUTHORS    Maurer,K. and Hillen,W.
TITLE       NEW BETA-GLUCANASE FROM A BACILLUS
JOURNAL    Patent: WO 9906573-A 2 11-FEB-1999;
MAURER KARL HEINZ (DE); HENKEL K&A (DE)
FEATURES
source
1..927
/organism="unidentified"
/strain="BACILLUS ALKALOPHILUS DSM 9956"
/db_xref="taxon:32644"

BASE COUNT      343 a 150 c 218 g 216 t

ORIGIN

Query Match           5.9%: Score 47.4; DB 9; Length 927;
Best Local Similarity 50.7%; Pred. No. 0.043;
Matches 173; Conservative 0; Mismatches 156; Indels 12; Gaps 2;

Oy      16   gatttaacggtgcccaactaacgcttagaagaagtacgctaagtttgaagcc 75

```

Db 286 GAATATAGCTGGGATTAAGACCAATCAGTTTATATCAATACGGGTTGTTGAAGTC 345
 QY 76 cgtatgaagaatgacgcagcagtcggagacagtgatccatgttccctccccaagatgt 135
 Db 346 AATATGAGCCAGCAATACAGGACCGCTCTTACTCTTTACATATACGGGTC 405
 QY 136 tccgaataatcgcagatggaagccctggtagaagtgatataatgaatcgcgcaagaat 135
 Db 406 TGGGATTGGATATG---ATCCTTGGGATGAATGATGATGATGATGATGATGATGAT 452
 QY 196 cgggagcgttccagtcacatcaatcaatcaatcaatcaatcaatcaatcaatcaat 255
 Db 463 ACAACAAGACTCCAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 513
 QY 256 aagcagcagtcgttagcccgccgagcagcagcagcagcagcagcagcagcagcagcag 315
 Db 514 CATTACACCAATTAAGGCTTCGATGCAATCAATCTTTTAAATACGATGCTTTGAATGG 573
 QY 316 actcgaatcagtcgcgctgagcagtcgtgacgagcagcagcagcagcagcagcagc 356
 Db 574 AGACCAAGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 614

RESULT 11
 ASTCELL90 306 bp DNA SYN 08-NOV-1996
 LOCUS Artificial sequence DNA for T-cell receptor, toolbox 90.
 DEFINITION Y09275
 ACCESSION Y09275
 VERSION Y09275.1 GI:1666290
 KEYWORDS T-cell receptor; toolbox 90.
 SOURCE synthetic construct.
 ORGANISM artificial sequence.
 REFERENCE 1 (bases 1 to 306)
 AUTHORS Schluesener, H.J.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 306)
 AUTHORS Schluesener, H.J.
 TITLE Direct Submission
 JOURNAL Submitted (21-OCT-1996) H.J. Schluesener, Inst.f. Hirnforschung, Calver Str. 3, D-77076 Tuebingen, FRG

FEATURES
 source
 1..306
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 1..306
 /note="toolbox 90"
 /codon_start=1
 /transl_table=11
 /product="T cell receptor peptide"
 /protein_id="CAA70470.1"
 /db_xref="GI:1666291"
 /translation="MASMTGGGQMGKRGSEFAADPGTIOQLFMSFNNQSLVELDM
 GHGLRLIHVSYSGSEFENGVDMDHGLRLIHVSYSNTEKLEKIAALEHHHHH"
 BASE COUNT 65 a 103 c 86 g 52 t
 ORIGIN

Query Match 5.7%; Score 45.8; DB 56; Length 306;
 Best Local Similarity 95.9%; Pred. No. 0.12;
 Matches 47; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 756 cgtgacaagaatgctggcgagcagcagcagcagcagcagcagcagcagcagcagcag 804
 Db 258 CCTCGAAGAGCTTGGCGCGCACTCGAGCACCACACACACACACACTGA 306

RESULT 12
 SYHAOM 1228 bp DNA SYN 04-APR-2000
 LOCUS Synthetic B.macerans/B.amyloliquefaciens hybrid gene for
 DEFINITION endo-1,3-1,4-beta-glucanase.

ACCESSION 225873
 VERSION 225873.1 GI:398185
 KEYWORDS beta-glucanase; endo-1,3-1,4-beta-glucanase.
 SOURCE synthetic construct.
 ORGANISM synthetic construct.
 REFERENCE 1 (bases 1 to 543)
 AUTHORS Hofemeister, J., Kurtz, A., Borris, R. and Knowles, J.
 TITLE The beta-glucanase gene from *Bacillus amyloliquefaciens* shows extensive homology with that of *Bacillus subtilis*
 JOURNAL Gene 49 (2), 177-187 (1986)
 MEDLINE 87192007
 REFERENCE 2 (bases 168 to 852)
 AUTHORS Borris, R., Buetner, K. and Maentzsaalae, P.
 TITLE Structure of the beta-1,3-1,4-glucanase gene of *Bacillus macerans*: homologues to other beta-glucanases
 JOURNAL Mol. Gen. Genet. 222 (2-3), 278-283 (1990)
 MEDLINE 91109712
 REFERENCE 3 (bases 1 to 1228)
 AUTHORS Pollitz, O., Simon, O., Olsen, O. and Borris, R.
 TITLE Determinants for the enhanced thermostability of hybrid (1-3,1-4)-beta-glucanases
 JOURNAL Eur. J. Biochem. 216 (3), 829-834 (1993)
 MEDLINE 94009045
 REFERENCE 4 (bases 1 to 1228)
 AUTHORS Pollitz, O., Simon, O., Olsen, O. and Borris, R.
 TITLE Determinants for the enhanced thermostability of hybrid (1-3,1-4)-beta-glucanases
 JOURNAL Eur. J. Biochem. (1993) In press
 REMARK 5 (bases 1 to 1228)
 REFERENCE Pollitz, O.
 TITLE Direct Submission
 JOURNAL Submitted (26-AUG-1993) Pollitz O., Humboldt University of Berlin, Biology, Warschauer Str. 43, BERLIN, Germany, D-10243

FEATURES
 source
 1..1228
 /organism="Bacillus amyloliquefaciens"
 /db_xref="taxon:1390"
 1..543
 /organism="Bacillus amyloliquefaciens"
 /db_xref="taxon:1390"
 1..1228
 /organism="Paenibacillus macerans"
 /db_xref="taxon:44252"
 469..543
 /function="endo-1,3-1,4-beta-glucanase"
 /note="hybrid between bgia (Pos. 1 to 543) and bgim (Pos. in hybrid 544 to 1228, Pos. original 168 to 852)"
 /codon_start=1
 /transl_table=11
 /label="HAOM
 /evidence="experimental
 /product="hybrid endo-1,3-1,4-beta-glucanase"
 /protein_id="CAA81092.1"
 /db_xref="GI:398186"
 /translation="MKRVLLIVGLFMSLCGITSVSAGSVFMEPLSYFNPSTWEKA
 DYSNMGVENCWTRANNVNFNDGKIKGLISLVNKRDCAEVRENTYNGVGLKYSIG
 KPAKNTGIVSSFFETFGARHGTOMDEIDIEFGKDTTVQGNRYTNGVGHKYSISG
 FDSKGFHTVAFDWPQGYIKWYVDGLVLTATNPSPGKIMNMLMNGTVDWGLGS
 YNGANPLAEYDWMKYTSN"
 544..1228
 /organism="Paenibacillus macerans"
 /db_xref="taxon:44252"
 371 a 220 c 290 g 347 t
 BASE COUNT
 ORIGIN

source

BASE COUNT 371 a 220 c 290 g 347 t
 ORIGIN

[illegible]

RESULT	13
SYHA12Y13-LOCUS	
DEFINITION	SYHA12Y13 1234 bp DNA SYN 04-APR-2000
ACCESSION	Synthetic B.macerans/B.amyloliquefaciens hybrid gene for endo-1,3-l,4-beta-glucanase.
VERSION	225876
KEYWORDS	endo-1,3-l,4-beta-glucanase.
SOURCE	225876.1 GI:398190
ORGANISM	beta-glucanase::endo-1,3-l,4-beta-glucanase.
	synthetic construct.
	synthetic construct.
	artificial sequence.
REFERENCE	1 (bases 1 to 579)
AUTHORS	Hofemeister,J., Kurtz,A., Borris,R. and Knowles,V.
TITLE	The beta-glucanase gene from Bacillus amyloliquefaciens shows extensive homology with that of Bacillus subtilis
JOURNAL	Gene 49 (2), 177-187 (1986)
MEDLINE	87192007
REFERENCE	2 (bases 198 to 852)
AUTHORS	Borris,R., Buettner,K. and Maentsaelae,P.
TITLE	Structure of the beta-1,3-l,4-gluconase gene of Bacillus macerans homologues to other beta-glucanases
JOURNAL	Mol. Gen. genet. 222 (2-3), 278-283 (1990)
MEDLINE	91109712
REFERENCE	3 (bases 1 to 1234)
AUTHORS	Poltz,O., Simon,O., Olsen,O. and Borris,R.
TITLE	Determinants for the enhanced thermostability of hybrid (1-3,l-4)-beta-glucanases
JOURNAL	Eur. J. Biochem. 216 (3), 829-834 (1993)
MEDLINE	9409045
REFERENCE	4 (bases 1 to 1234)
AUTHORS	Poltz,O., Simon,O., Olsen,O. and Borris,R.
TITLE	Determinants for the enhanced thermostability of hybrid (1-3,l-4)-beta-glucanases
JOURNAL	Eur. J. Biochem. (1993) In press
REMARK	(sites)
REFERENCE	5 (bases 1 to 1234)
AUTHORS	Poltz,O.
TITLE	Direct Submission
JOURNAL	Submitted (26-AUG-1993) Poltitz O., Humboldt University of Berlin, Biology, Marschner Str. 43, BERLIN, Germany, D-10243
FEATURES	Location/Qualifiers
source	1..1234
	/organism="synthetic construct"
	/db_xref="taxon:32630"
	/focuss
	1..1234
source	/organism="Bacillus amyloliquefaciens"
	/db_xref="taxon:1390"

```

source
1. .579
/organism="Bacillus amylioliquefaciens"
/db_xref="taxon:1390"
1. .1234
/organism="Paenibacillus macerans"
/db_xref="taxon:44252"
469. .543
469. .1188
/function="endo-1,3-1,4-beta-glucanase"
/notes="Hybrid between bglA (Pos. 1 to 579) and bglM
(Pos. in hybrid 580 to 1234, Pos. original 198 to 852)
Deletion of Y13 of HA12-M"
/codon_start=1
/transl_table=1
/label=HA12M
/evidence=experimental
/product="hybrid-endo-1,3-1,4-beta-glucanase"
/protein_id="CA81095.1"
/db_xref="GI:398191"
/translacion="MKRRVLLIVYGLFMSLCGTTSSVATQGSFFEPFSPNSTRE
KADYSSNGVFNCTIRANWNFTNDGKLGKLGITSSAVNKFDCAEYRSTNIYGGLEEV
SMKPAKGTGTSSEFFYTGPAHGTOIMEIDIEFGKPTKVOFNYYNGVGGEHKYIS
LGFEAGKGFHYAFDMQGPYIKWYVDVCLKHTATANIPTIPGKIMNMLNMGTDGVDWL
GSYGANPLVAEYIMWKYTSN"
580. .1234
/organism="Paenibacillus macerans"
/db_xref="taxon:44252"

```

	Query Match	5.5%;	Score 44.6;	DB 56;	Length 1234;	
	Best Local Similarity	52.9%;	Pred. No. 0.24;	Mismatches 154;	Conservative 0;	Mismatches 119; Indels 18; Gaps 2;
Oy	58	tacggaagtttaagaccgatatgaagtggcaagccgcatacggagaacagtcagttccatg	117			
Dd	757	TACGGCCCTGTACGAGGTGATGAAGGCCACCACCAAAAAATACAGGAATTGTTCACCTTTT	816			
Oy	118	ttcccctcacgaataggtctccgaattcgcatgcgaatgaagcccttggtataagtgatat	177			
Dd	817	TTCACGTA-----TACACGACCTGCTCATGGCACAACATGGAGTAATAGATATTC	867			
Oy	178	gaagttcttcggcaagaatccgagccagttccaglttccaacatcatcaccggaatgaagccgyc	237			
Dd	868	GAAATTCTTAGGAAAAGACACGCACAAAAGTCCAGTTACTATTATACCAATGGGGTTTGGC	927			
Oy	238	gcacaaagactagaggaaaagacacatctgttgtgccccgcgccgatacagttcttcac	297			
Dd	928	GCTCAT-----GAAAGAGGTATATCTCTTGCGCTTTGATGTCACAAAGGCTTCCTAC	978			
Oy	298	aacctcaggtctcgaatgacgtccgaattacgltccgcgtgacgtgttagcagct	348			
Dd	979	ACCTATGCTTTGATTTGGCAGCGCAGGATATTTAATGATGTAATGACGCT	1029			
RESULT	14					
SYHA16M						
LOCUS	1234 bp	DNA	SVN	04-APR-2000		
DEFINITION	Synthetic B.macerans/B.amyloliquefaciens hybrid gene for endo-1,3-1,4-beta-glucanase.					
ACCESSION	Z25878					
VERSION	Z25878.1	GI:398194				
KEYWORDS	beta-glucanase; endo-1,3-1,4-beta-glucanase.					
SOURCE	synthetic construct.					
ORGANISM	artificial sequence.					
REFERENCE	1 (bases 1 to 591)					
AUTHORS	Hofemeister,J., Kurtz,A., Borriis,R. and Knowles,J.					
TITLE	The beta-glucanase gene from Bacillus amyloliquefaciens shows extensive homology with that of Bacillus subtilis					
JOURNAL	Gene 49 (2), 177-187 (1986)					
MEDLINE	87192007					

Oy	178	gaactctcgcaagaatccggcagttccacatccaatcatcgcgtaaggccgcg	237
Dd	868	GAATTCTAGGAAAGACGACGACAAAAGTCATTTTAATACTATTATACCAATGTGGTTCGC	927
Oy	238	gcacaaagactgcgaaaagcaccatgtcgttagccccgcgcgcgatcagcttccac	297
Dd	928	GGTCTAT-----GAAAGGTTATCTCCTCTTGCGTTTGATGATCATCAACAGGCTTCAT	978
Oy	298	accatcagcttcgaatgacctccgaattacgttcgcttgactggttacgagt	348
Dd	979	ACCTATCCTTTTCGATTGTGGACGACGAGGTATATTAAAAAGTAGTACGCGT	1029
RESULT	15		
LOCUS	SYHA2M	1234 bp	DNA SYN 04-APR-2000
DEFINITION	Synthetic B.macerans/B.amyloliquefaciens hybrid gene for endo-1,-3-1,4-beta-glucanase.		
ACCESSION	225879		
VERSION	225879.1	GI:398196	
KEYWORDS	beta-glucanase; endo-1,3-1,4-beta-glucanase.		
SOURCE	synthetic construct. artificial sequence.		
ORGANISM	1 (bases 1 to 549) Hofemeister,J., Kurte,A., Borrijs,R. and Knowles,J. The beta-glucanase gene from Bacillus amyloliquefaciens shows extensive homology with that of Bacillus subtilis Gene 49 (2), 177-187 (1986)		
REFERENCE	2 (bases 168 to 852) Borrijs,R., Buettner,K. and Maentzsaetiae,P. Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans homologies to other beta-glucanases Mol. Gen. genet. 222 (2-3), 278-283 (1990)		
AUTHORS	3 (bases 1 to 1234) Politz,O., Simon,O., Olsen,O. and Borrijs,R. Determinants for the enhanced thermostability of hybrid (1-3,1-4)-beta-glucanases Eur. J. Biochem. 216 (3), 829-834 (1993)		
JOURNAL	MEDLINE 94009045		
REFERENCE	4 (bases 1 to 1234) Politz,O., Simon,O., Olsen,O. and Borrijs,R. Determinants for the enhanced thermostability of hybrid (1-3,1-4)-beta-glucanases Eur. J. Biochem. (1993) In press		
AUTHORS	JOURNAL REMARK REFERENCE AUTHORS	5 (bases 1 to 1234) Politz,O. Submitted Direct Submission Submitted (26-AUG-1993) Politz O., Humboldt University of Berlin, Biology, Warschauer Str. 43, BERLIN, Germany, D-10243	
TITLE	JOURNAL TITLE	Location/Qualifiers 1. .1234 /organism="synthetic construct" /db_xref="taxon:32630"	
JOURNAL	source	/focus 1. .1234 /organism="Bacillus amyloliquefaciens" /db_xref="taxon:1390"	
FEATURES	source	1. .549 /organism="Bacillus amyloliquefaciens" /db_xref="taxon:1390" 1. .1234 /organism="Paenibacillus macerans" /db_xref="taxon:44252" 469. .543 469. .1188 /function="endo-1,3-1,4-beta-glucanase" /note="hybrid between bsla (Pos. 1 to 549) and bglm (pos.in hybrid 550 to 1334, Pos. original 168 to 852)"	
source	sig_peptide	CDS	

```
/codon_start=1
/transl_table=11
/label=HA2M
/evidence=experimental
/product=hybrid-endo-1,3-1,4-beta-glucanase"
/protein_id="CAA81098.1"
/db_xref="GI:398197"
/translation="MKRVLLILVTGLFMSICGITSVSQAQGSVFMEPLSYFNPSTWE
KADGYSNGVFNCTWRANNVNFINDGKRLGLTSSAVNKFDCAEYRSTNYGLEYEY
SMKPAKNTGIVSSFFTYTGPAGHGTOMDEIDIEFLGKDTTFKQFNYNTNGVGHKVIS
LGPDASKGFHTTYAFDMQPGYIKRYVDGVLLKHTATANIPTSPGKIMANLNGTGVDDML
GSYNGANPLTYAETDWMVKYTSN"
550..1234
/organism="Paenibacillus macerans"
/db_xref="taxon:44252"
```

BASE COUNT 375 a 222 c 290 g 347 t
ORIGIN

Query Match 5.5%; Score 44.6; DB 56; Length 1234;
Best Local Similarity 52.9%; Pred. No. 0.24;
Matches 154; Conservative 0; Mismatches 119; Indels 18; Gaps 2;

```
OY 58 taaggtaagttgaagccgtaigaagatgagccgcatcggaacagtcagttcatag 117
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 757 TACGGCCTGTACGAGGTGATGATGAGCCACCAAAAATACAGGAATTGTCTCATCCTT 816

OY 118 ttccctaccagaatggttcggaatcgccgattggaagccctggttagaagtgaatt 177
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 817 TTCAAGTA-----TACAGGACCTGCTCATGTGCACACAAATGGGATGAATATATC 867

OY 178 gaagttctcgcaagaatccggcagtttcagttccaacatcaltaccgylaagccggc 237
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 868 GAATTTCTAGGAAAGACACACAAAGTTCAGTTTAACTTATATACCAATGGGTTGCG 927

OY 238 gcacaaaagactagcgaaaaagcaccatgctgttagcccgccgacgaltcaggcttcac 297
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 928 GGTGAT-----GAAAAGGTTATCTCTCTGTGCTTTGATGCATCAAGGGCTTCCAT 978

OY 298 acctagcgttcgaatgactccgaattacgtccgctggaactgttgacggt 348
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 979 ACCTATGCTTTCGATTTGGACGCCAGGATATTAATGTATGTATGACGGT 1029
```

Search completed: August 23, 2001, 13:37:38
Job time: 8301 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 12:16:57 ; Search time 157.43 Seconds
(without alignments)
3206.713 Million cell updates/sec

Title: US-09-654-652a-5

Perfect score: 804
1 atcgttagcgcacaagattt.....accacacacacacacactga 804

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601:*

1: /cgnl_9/gcgdata/geneseq/geneseqn/NA1980.DAT:*

2: /cgnl_9/gcgdata/geneseq/geneseqn/NA1981.DAT:*

3: /cgnl_9/gcgdata/geneseq/geneseqn/NA1982.DAT:*

4: /cgnl_9/gcgdata/geneseq/geneseqn/NA1983.DAT:*

5: /cgnl_9/gcgdata/geneseq/geneseqn/NA1984.DAT:*

6: /cgnl_9/gcgdata/geneseq/geneseqn/NA1985.DAT:*

7: /cgnl_9/gcgdata/geneseq/geneseqn/NA1986.DAT:*

8: /cgnl_9/gcgdata/geneseq/geneseqn/NA1987.DAT:*

9: /cgnl_9/gcgdata/geneseq/geneseqn/NA1988.DAT:*

10: /cgnl_9/gcgdata/geneseq/geneseqn/NA1989.DAT:*

11: /cgnl_9/gcgdata/geneseq/geneseqn/NA1990.DAT:*

12: /cgnl_9/gcgdata/geneseq/geneseqn/NA1991.DAT:*

13: /cgnl_9/gcgdata/geneseq/geneseqn/NA1992.DAT:*

14: /cgnl_9/gcgdata/geneseq/geneseqn/NA1993.DAT:*

15: /cgnl_9/gcgdata/geneseq/geneseqn/NA1994.DAT:*

16: /cgnl_9/gcgdata/geneseq/geneseqn/NA1995.DAT:*

17: /cgnl_9/gcgdata/geneseq/geneseqn/NA1996.DAT:*

18: /cgnl_9/gcgdata/geneseq/geneseqn/NA1997.DAT:*

19: /cgnl_9/gcgdata/geneseq/geneseqn/NA1998.DAT:*

20: /cgnl_9/gcgdata/geneseq/geneseqn/NA1999.DAT:*

21: /cgnl_9/gcgdata/geneseq/geneseqn/NA2000.DAT:*

22: /cgnl_9/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	65.6	8.2	5248 16	AAAT08489
2	65.6	8.2	5248 17	AAAT1898
3	64.8	8.1	5248 20	AAAT2808
4	64.8	8.1	5248 20	AAAT2808
5	61.4	7.6	2735 17	AAAT4351
6	61.4	7.6	2007 20	AAAT80369
7	61.4	7.6	2007 20	AAAT80370
8	61.4	7.6	2007 20	AAAT80366
9	61.4	7.6	2007 20	AAAT80367
10	61.4	7.6	2007 20	AAAT80368
11	55.5	6.8	6363 20	AAAT80368

12	50.8	6.3	59	22	AAAT74635
13	47.4	5.9	927	20	AAAT02912
14	44.6	5.5	783	11	AAAT05167
15	44.6	5.5	850	11	AAAT03519
16	42.6	5.3	586	21	AAAT61166
17	42.4	5.3	423	20	AAAT19396
18	42.4	5.3	423	21	AAAT19184
19	42.2	5.2	1240	11	AAAT05832
20	42.2	5.2	357	21	AAAT05832
21	42.2	5.2	1186	20	AAAT61167
22	42.2	5.2	1247	20	AAAT9725
23	42.2	5.2	6619	20	AAAT3789
24	41.6	5.2	550	21	AAAT1987
25	38.8	4.8	1815	19	AAAT39341
26	38.8	4.8	1815	19	AAAT39341
27	37.6	4.7	1166	20	AAAT19440
28	37.6	4.7	1166	20	AAAT19440
29	37.4	4.7	1166	20	AAAT19440
30	37.2	4.6	607	21	AAAT4807
31	35.6	4.4	1029	22	AAAT60750
32	35.2	4.4	445	21	AAAT1502
33	35.2	4.4	1195	20	AAAT19453
34	35.2	4.4	1195	20	AAAT19453
35	34.6	4.3	556	21	AAAT09632
36	34.2	4.3	948	16	AAAT08748
37	34.2	4.3	948	17	AAAT27687
38	34.2	4.3	948	18	AAAT70349
39	34.2	4.3	948	18	AAAT76649
40	34.2	4.3	948	19	AAAT65799
41	34.2	4.3	948	19	AAAT65799
42	34.2	4.3	948	20	AAAT63425
43	34.2	4.3	963	18	AAAT76626
44	34.2	4.3	963	19	AAAT65831
45	34.2	4.3	963	19	AAAT53941

ALIGNMENTS

RESULT 1	
AAAT08489	
ID	AAAT08489 standard; DNA; 5248 BP.
AC	AAAT08489;
XX	
DT	15-MAR-1996 (first entry)
XX	
DE	Plasmid PET TRC S05/N1 containing 561.1 scFv.
XX	
KW	Complement C5; haemolysis; kidney; glomerulonephritis;
KW	monoclonal antibody; antiinflammatory; antibody engineering;
KW	humanised antibody; complementarity determining region; CDR;
KW	single chain antibody; scFv; vector; PET TRC S05/N1;
KW	Escherichia coli; ds.
XX	
OS	Synthetic.
XX	
PN	W09529697-A1.
XX	
PD	09-NOV-1995.
XX	
PF	01-MAY-1995; 95WO-US05688.
XX	
PR	02-MAY-1994; 94US-0236208.
XX	
PA	(ALEX-) ALEXION PHARM INC.
XX	
PI	Evans MJ, Matlis L, Mueller EE, Nye SH, Rollins S;
PI	Rother RP, Springhorn J P, Squinto SP, Thomas TC;
XX	Wang Y, Wilkins JA;
DR	WPI; 1995-392923/50.
XX	

Cloning vector pRK
B. alkalophilus be
Heat-stable endo-b
Thermostable beta-
SEN virus genome f
M. tuberculosis an
Beta-glucanase-H1
SEN virus genome f
Insert from pET28C
Insert from pET28C
Human VEGF-X CUB-1
Human papillomavir
Human papillomavir
M. tuberculosis an
M. tuberculosis re
C. trachomatis clo
H. contortus clone
Neocarzinostatin a
M. tuberculosis re
Fusarium venenatum
5' Nuclease from T
Mutant Thermus aqu
Synthesis deficient
5' Nuclease Cleava
Thermus aquaticus
Nucleotide sequenc
DNA sequence of a
Cleavease BN/thromb
Cleavease BN/thromb
Nucleotide sequenc

CC clones.
 XX
 SQ Sequence 4411 BP; 1065 A; 1156 C; 1124 G; 1066 T; 0 other;
 Query Match 8.1%; Score 64.8; DB 20; Length 4411;
 Best Local Similarity 97.1%; Pred. No. 1.4e-09;
 Matches 66; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Db 737 ctgtccgaatcgaagctccgtcgcacaaagcttgcgcgcacatcgacacaccaccacc 796
 1719 CGGATTCGAATCGAGCTCCGTGACACAAGCTTGGCGCCGACACTGACACACCAACCAACC 1660
 QY 797 accactga 804
 Db 1659 ACCACTGA 1652

RESULT 4
 AAX28085
 ID AAX28085 standard; DNA; 5616 BP.
 XX
 AC AAX28085;
 DT 10-JUN-1999 (first entry)
 XX
 DE Plasmid pET21.tac.
 XX
 KM MLV reverse transcriptase; dscDNA production; cDNA library production;
 KW ss.
 XX
 OS Synthetic.
 OS
 PN US5891637-A.
 XX
 PD 06-APR-1999.
 XX
 PF 15-SEP-1997; 9705-0929967.
 XX
 PR 15-OCT-1996; 9605-0732861.
 PR 15-SEP-1997; 9705-0929967.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ruppert SJ;
 XX
 DR WPI; 1999-253852/21.
 XX
 PT Producing double stranded cDNA molecules from mRNA transcripts and a
 PT novel xenotropic murine leukemia virus reverse transcriptase -
 PT useful for constructing cDNA libraries
 XX
 PS Example 5; Fig 7; 70pp; English.
 XX
 CC This sequence represents a plasmid used to test the method of
 CC the invention. The method is for the production of double
 CC stranded cDNA (dscDNA) molecules from mRNA transcripts and a xenotropic
 CC murine leukemia virus reverse transcriptase (XM-MLV). The method, the
 CC host cells, reverse transcriptases and nucleic acids disclosed may be
 CC used to produce high quality, normalized, full-length, directionally
 CC cloned, cDNA libraries. Use of the method to construct cDNA libraries
 CC ensures that the reverse transcriptase is able to extend the first strand of
 CC the cDNA up to the ultimate 5' end of the mRNA, therefore ensuring that
 CC all sequences are represented in the library, whereas previously, some
 CC sequences would be lost due to incomplete transcription (e.g. as a result
 CC of hairpin loop formation) or mRNA degradation (e.g. as a result of RNase
 CC H activity). Also, carrying the process out in cells means that enzymes
 CC and proteins present in the cell will repair any mistakes or nicks in the
 CC product DNA. Modifying the mRNA for insertion into a vector by adding the
 CC 5' cap provides a convenient way of priming it for cDNA synthesis. Also,
 CC the method provides a fast and clean way of synthesizing cDNA
 CC clones.
 XX

SQ Sequence 5616 BP; 1306 A; 1478 C; 1492 G; 1340 T; 0 other;
 Query Match 8.1%; Score 64.8; DB 20; Length 5616;
 Best Local Similarity 97.1%; Pred. No. 1.5e-09;
 Matches 66; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Db 737 ctgtccgaatcgaagctccgtcgcacaaagcttgcgcgcacatcgacacaccaccacc 796
 5413 cgtgtccgaatcgaagctccgtcgcacaaagcttgcgcgcacatcgacacaccaccacc 5472
 QY 797 accactga 804
 Db 5473 accactga 5480

RESULT 5
 AAT44351
 ID AAT44351 standard; cDNA; 2735 BP.
 XX
 AC AAT44351;
 XX
 DT 11-FEB-1997 (first entry)
 XX
 DE H. pylori ureA + ureB gene locus in pORV214.
 XX
 KM Urease; urea gene; ureB gene; vaccine; ds.
 XX
 OS Chimeric Helicobacter strain CPM630;
 OS Chimeric bacteriophage T7.
 XX
 FH Key
 FH Location/Qualifiers
 FT 1..16
 FT promoter
 FT /**tag= a
 FT /note= "T7 promoter provides transcription
 FT initiation for the urease genes"
 FT misc_signal
 FT 33..43
 FT /**tag= b
 FT /function= operator
 FT /note= "lac operator provides inducible expression
 FT of the urease genes"
 FT primer_bind
 FT complement (46..67)
 FT /**tag= c
 FT /note= "BL1 primer"
 FT 102..818
 FT /**tag= d
 FT /product= urease A subunit
 FT CDS
 FT 822..2531
 FT /**tag= e
 FT /product= urease B subunit
 FT primer_bind
 FT 2546..2569
 FT /**tag= f
 FT /note= "BL2 primer"
 FT 2693..2735
 FT /**tag= g
 FT /note= "T7 terminator"
 FT terminator
 FT W09633732-A1.
 PN
 XX
 PD 31-OCT-1996.
 XX
 PF 25-APR-1996; 96WO-US05800.
 XX
 PR 06-DEC-1995; 95US-0568122.
 PR 28-APR-1995; 95US-0431041.
 XX
 PA (ORAV-) ORAVAX INC.
 XX
 PI Ackerman SK, Bhagat H, Ermak T, Guirakhoo F, Kleantous H;
 PI Lee CK, Monath TP, Pappo J, Soman G, Sussman I;
 PI Thomas WD, Weltzin RA.
 XX
 DR WPI; 1996-497373/49.

DR P-PSDB; AAM07193;
DR P-PSDB; AAM07194.
XX Vaccine for inducing mucosal response to Helicobacter - contg.
PT multimeric urease complex and pref. an antibiotic, anti-secretory
PT agent or bismuth salt
XX
XX
XX Disclosure; Page 68-70; 98pp; English.
PS
XX A cDNA clone (AAT44351), derived from PORV214, includes the urea and
CC urea sequences coding for the urease A (AAM07193) and urease B
CC (AAM07194) subunits of Helicobacter pylori clinical isolate CPM630.
CC To obtain PORV214, a genomic DNA library of CPM630 was screened
CC with anti-Helicobacter urease antibody. A 17 kb SalI fragment from
CC an isolated clone was subcloned into pUC18 to give pSCP1. PCR
CC primers (AAT44352-53) were used to amplify a 2.5 kb fragment from
CC pSCP1, which was inserted into pET24+ to give PORV214. The vector
CC was utilized in the prodn. of recombinant, enzymatically inactive,
CC multimeric urease in E. coli transformants for use in vaccines to
CC treat or prevent Helicobacter infection.
XX
SQ Sequence 2735 BP; 875 A; 562 C; 635 G; 663 T; 0 other;

Query Match 7.6%; Score 61.4; DB 17; Length 2735;
Best Local Similarity 98.4%; Pred. No. 1.2e-08;
Matches 62; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 742 cgaattcgagctccgtcgacaaagcttgaggcgccgactcgagaccacccaccacac 801
DB 2562 cgaattcgagctccgtcgacaaagcttgaggcgccgactcgagaccacccaccac 2621
OY 802 tga 804
DB 2622 tga 2624

RESULT 6
AAX80369
ID AAX80369 standard; cDNA; 2007 BP.
XX
AC AAX80369;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:119.
XX
KM HCV; hepatitis C virus; single chain recombinant complex; linker;
KM NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KM hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO9928482-A2.
XX
PD 10-JUN-1999.
XX
PF 24-NOV-1998; 98WO-US24528.
XX
PR 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX
PA (SCHE) SCHERING CORP.
PI Malcolm BA, Taremi SS, Weber PC, Yao N;
DR WPI: 1999-385385/32.
XX
XX New hepatitis C virus covalent complexes
XX
PS Disclosure; Page 204-207; 211pp; English.
XX

CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX
SQ Sequence 2007 BP; 413 A; 621 C; 563 G; 410 T; 0 other;

Query Match 7.6%; Score 61; DB 20; Length 2007;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 gaattcgagctccgtcgacaaagcttgaggcgccgactcgagaccacccaccacac 803
DB 1942 gaattcgagctccgtcgacaaagcttgaggcgccgactcgagaccacccaccacac 2001
OY 804 a 804
DB 2002 a 2002

RESULT 7
AAX80370
ID AAX80370 standard; cDNA; 2007 BP.
XX
AC AAX80370;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:120.
XX
KM HCV; hepatitis C virus; single chain recombinant complex; linker;
KM NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KM hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO9928482-A2.
XX
PD 10-JUN-1999.
XX
PF 24-NOV-1998; 98WO-US24528.
XX
PR 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX
PA (SCHE) SCHERING CORP.
PI Malcolm BA, Taremi SS, Weber PC, Yao N;
DR WPI: 1999-385385/32.
XX
XX New hepatitis C virus covalent complexes
XX
PS Disclosure; Page 207-210; 211pp; English.

CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.

CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.

XX
SQ Sequence 2007 BP; 413 A; 622 C; 563 G; 409 T; 0 other;

Query Match 7.6%; Score 61; DB 20; Length 2007;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 744 gaattcgagctcgcgtgacaaagcttgcgcgcactcgcagcaccaccaccaccactg 803
|||||

DB 1942 gaattcgagctcgcgtgacaaagcttgcgcgcactcgcagcaccaccaccaccactg 2001

OY 804 a 804
DB 2002 a 2002

RESULT 8

AAx80366 standard; cDNA; 2007 BP.

XX AAX80366;

XX 07-SEP-1999 (first entry)

DE HCV NS4A-NS3 complex encoding cDNA SEQ. ID NO:116.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
KM NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KM hydrophobic domain; covalent complex; detection; inhibitor; ss.

XX Hepatitis C virus.

OS Synthetic.

PN MO928482-A2.

PD 10-JUN-1999.

PE 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

PR 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.

PI Malcolm BA, Taremi SS, Weber PC, Yao N;

DR WPI; 1999-385385/32.

XX New hepatitis C virus covalent complexes

PS Disclosure; Page 194-197; 211pp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.

SQ Sequence 2007 BP; 415 A; 619 C; 564 G; 409 T; 0 other;

Query Match 7.6%; Score 61; DB 20; Length 2007;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 gaattcgagctcgcgtgacaaagcttgcgcgcactcgcagcaccaccaccaccactg 803
|||||

DB 1942 gaattcgagctcgcgtgacaaagcttgcgcgcactcgcagcaccaccaccaccactg 2001

OY 804 a 804
DB 2002 a 2002

RESULT 9

AAx80367 standard; cDNA; 2007 BP.

XX AAX80367;

XX 07-SEP-1999 (first entry)

DE HCV NS4A-NS3 complex encoding cDNA SEQ. ID NO:117.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;
KM NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KM hydrophobic domain; covalent complex; detection; inhibitor; ss.

XX Hepatitis C virus.

OS Synthetic.

PN MO928482-A2.

PD 10-JUN-1999.

PE 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

PR 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.

PI Malcolm BA, Taremi SS, Weber PC, Yao N;

DR WPI; 1999-385385/32.

XX New hepatitis C virus covalent complexes

PS Disclosure; Page 197-200; 211pp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.

SQ Sequence 2007 BP; 415 A; 619 C; 564 G; 409 T; 0 other;

Query Match 7.6%; Score 61; DB 20; Length 2007;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 gaattcgagctcgcgtgacaaagcttgcgcgcactcgcagcaccaccaccaccactg 803
|||||

DB 1942 gaattcgagctcgcgtgacaaagcttgcgcgcactcgcagcaccaccaccaccactg 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 12:48:29 ; Search time 1874.7 Seconds
(without alignments)
4054.032 Million cell updates/sec

Title: US-09-654-652A-5
Perfect score: 804
Sequence: 1 atggtagcgcaagaagattt.....accacacacacacactga 804

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1.*
2: gb_est12.*
3: gb_est3.*
4: gb_est14.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*

44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_esthum29.*
64: em_esthum30.*
65: em_esthum31.*
66: em_esthum32.*
67: em_esthum33.*
68: em_esthum34.*
69: em_esthum35.*
70: em_esthum36.*
71: em_esthum37.*
72: em_esthum38.*
73: em_esthum39.*
74: em_esthum40.*
75: em_esthum41.*
76: em_esthum42.*
77: em_esthum43.*
78: em_esthum44.*
79: em_esthum45.*
80: em_esthum46.*
81: em_esthum47.*
82: em_esthum48.*
83: em_esthum49.*
84: em_esthum50.*
85: em_esthum51.*
86: em_esthum52.*
87: em_esthum53.*
88: em_esthum54.*
89: em_esthum55.*
90: em_esthum56.*
91: em_esthum57.*
92: em_esthum58.*
93: em_esthum59.*
94: em_esthum60.*
95: em_esthum61.*
96: em_esthum62.*
97: em_esthum63.*
98: em_esthum64.*
99: em_esthum65.*
100: em_esthum66.*
101: em_esthum67.*
102: em_esthum68.*
103: em_esthum69.*
104: em_esthum70.*
105: em_esthum71.*
106: em_esthum72.*
107: em_esthum73.*
108: em_esthum74.*
109: em_esthum75.*
110: em_esthum76.*
111: em_esthum77.*
112: em_esthum78.*
113: em_esthum79.*
114: em_esthum80.*
115: em_esthum81.*
116: em_esthum82.*

117: gb_est148:*
118: gb_est149:*
119: gb_est150:*
120: gb_est151:*
121: gb_est152:*
122: gb_est153:*
123: gb_est154:*
124: gb_est155:*
125: gb_est156:*
126: gb_est157:*
127: gb_est158:*
128: gb_est159:*
129: gb_est160:*
130: gb_est161:*
131: gb_est162:*
132: gb_est163:*
133: gb_est164:*
134: gb_est165:*
135: gb_est166:*
136: gb_est167:*
137: gb_est168:*
138: gb_est169:*
139: gb_est170:*
140: gb_est171:*
141: gb_est172:*
142: gb_est173:*
143: gb_est174:*
144: gb_est175:*
145: gb_est176:*
146: gb_est177:*
147: gb_est178:*
148: gb_est179:*
149: gb_est180:*
150: gb_est181:*
151: gb_est182:*
152: gb_est183:*
153: gb_est184:*
154: gb_est185:*
155: gb_est186:*
156: gb_est187:*
157: gb_est188:*
158: gb_est189:*
159: gb_est190:*
160: gb_est191:*
161: gb_est192:*
162: gb_est193:*
163: gb_est194:*
164: gb_est195:*
165: gb_est196:*
166: gb_est197:*
167: gb_est198:*
168: gb_est199:*
169: gb_est200:*
170: gb_est201:*
171: gb_est202:*
172: gb_est203:*
173: gb_est204:*
174: gb_est205:*
175: gb_est206:*
176: gb_est207:*
177: gb_est208:*
178: gb_est209:*
179: gb_est210:*
180: gb_est211:*
181: gb_est212:*
182: gb_est213:*
183: gb_est214:*
184: gb_est215:*
185: gb_est216:*
186: gb_est217:*
187: gb_est218:*
188: gb_est219:*
189: gb_est220:*

190: gb_est110:*
191: gb_est111:*
192: gb_hrc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_inv4:*
207: em_gss_inv5:*
208: em_gss_inv6:*
209: em_gss_inv7:*
210: em_gss_inv8:*
211: em_gss_inv9:*
212: em_gss_inv10:*
213: em_gss_inv11:*
214: em_gss_inv12:*
215: em_gss_inv13:*
216: em_gss_inv14:*
217: em_gss_inv15:*
218: em_gss_inv16:*
219: em_gss_inv17:*
220: em_gss_inv18:*
221: em_gss_inv19:*
222: em_gss_inv20:*
223: em_gss_inv21:*
224: em_gss_inv22:*
225: em_gss_inv23:*
226: em_gss_inv24:*
227: em_gss_inv25:*
228: em_gss_inv26:*
229: em_gss_inv27:*
230: em_gss_inv28:*
231: em_gss_inv29:*
232: em_gss_inv30:*
233: em_gss_inv31:*
234: em_gss_inv32:*
235: em_gss_inv33:*
236: em_gss_inv34:*
237: em_gss_inv35:*
238: em_gss_inv36:*
239: em_gss_inv37:*
240: em_gss_inv38:*
241: em_gss_inv39:*
242: em_gss_inv40:*
243: em_gss_inv41:*
244: em_gss_inv42:*
245: em_gss_inv43:*
246: em_gss_inv44:*
247: em_gss_inv45:*
248: em_gss_inv46:*
249: em_gss_inv47:*
250: em_gss_inv48:*
251: em_gss_inv49:*
252: em_gss_inv50:*
253: em_gss_inv51:*
254: em_gss_inv52:*
255: em_gss_inv53:*
256: em_gss_inv54:*
257: em_gss_inv55:*
258: em_gss_inv56:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	47	5.8	599	155	BG605030	WHE2326-D
2	45.4	5.6	529	136	BE494282	WHE1253_H
3	41	5.1	418	136	BE495688	WHE1279_B
4	37.8	4.7	796	15	A1068991	mgae0005a
5	37.8	4.7	849	15	A1069170	mgae0005d
6	37.2	4.6	561	13	BG103002	RH122_35
7	37	4.6	552	19	A1397807	NCR9C4T3
8	36.8	4.6	747	170	BF661648	BF861648 963025E08
9	36.4	4.5	408	158	H92520	ys85a11.r1
10	36.4	4.5	604	166	BE359676	DC1_56_H0
11	36.4	4.5	608	154	BE488141	RH122_60
12	36.2	4.5	635	167	BE405426	WHE1216_C
13	35.8	4.5	387	145	BF202533	WHE1776_G
14	35.8	4.5	584	13	AA898069	NCM3G7V7
15	35.8	4.5	587	13	AA897960	NCM1H9T7
16	35.4	4.4	399	237	A2046873	sheb0088H
17	35.4	4.4	640	141	BE888160	601511671
18	35.4	4.4	669	164	BE208885	GF-EV-P3C
19	35.4	4.4	944	169	BF782886	602107758
20	35.2	4.4	360	167	BE428429	MTD006.H1
21	35.2	4.4	423	167	BE398687	WHE0024.C
22	35.2	4.4	536	149	BE473353	WHE0923.E
23	35.2	4.4	552	167	BE419465	WMS012.D8
24	35.2	4.4	776	151	BF627470	HVSMED000
25	35	4.4	340	116	AW486705	76371.MAR
26	35	4.4	482	166	BE347923	SP08H02.Y
27	35	4.4	495	166	BE059962	sn39C06.Y
28	35	4.4	649	230	AO575231	nxbx0087A
29	35	4.4	762	170	BF665955	963063D08
30	35	4.4	781	141	BE884266	601505733
31	34.8	4.3	826	146	BF263101	HV-CEA000
32	34.6	4.3	876	150	BF526437	602070921
33	34.4	4.3	443	104	A1940919	SB18111.Y
34	34.4	4.3	474	33	AV668531	NCM6G5T7
35	34.2	4.3	376	144	BF072772	NCM5G11T
36	34.2	4.3	455	175	BE278686	ade05np.f
37	34.2	4.3	501	175	BE279070	adf05np.f
38	34.2	4.3	521	175	BE279385	b2g07np.f
39	34.2	4.3	532	175	BE279532	b3g11np.f
40	34.2	4.3	546	30	AV386476	AV386476 AV386476
41	34.2	4.3	548	13	AA898699	NCM6G5T7
42	34.2	4.3	550	13	AA898546	NCP5G12T7
43	34.2	4.3	552	175	BE279542	b3h07np.f
44	34.2	4.3	558	13	AA898246	NCM4A9T7
45	34.2	4.3	560	175	BE280207	c1f11np.f

ALIGNMENTS

RESULT 1
 BG605030 599 bp mRNA EST 16-APR-2001
 LOCUS WHE2326_D09.G1875 wheat pre-anthesis spike cDNA library Triticum
 DEFINITION aestivum cDNA clone WHE2326_D09.G18, mRNA sequence.
 ACCESSION BG605030
 VERSION BG605030.1 GI:13635033
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Triticum.
 1 (bases 1 to 599)
 REFERENCE Anderson O.D., Chao S., Choi D.W., Close T.J., Fenton R.D., Han
 P.S., Hsiao C.C., Kang Y., Lazo G.R., Miller R., Rausch C.J.,
 Seaton C.L. and Tong J.C.

TITLE The structure and function of the expressed portion of the wheat
 JOURNAL genomes - Pre-anthesis spike cDNA library
 COMMENT Unpublished (2000)
 CONTACT Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 510595773
 Fax: 510595818
 Email: oanderson@wpr.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Stragene SK primer.
 Location/Qualifiers
 1..599
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone_lib="WHE2326_D09.G18"
 /clone_lib="Wheat pre-anthesis spike cDNA library"
 /tissue_type="Spike before anthesis"
 /dev_stage="Adult plant"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
 greenhouse. Whole spike with awns trimmed, white, green
 and yellow anther were collected and total RNA, and
 poly(A) RNA were prepared, a cDNA library was made, and
 the cDNA clones were in vivo excised to give phagescript
 phagemids in the T7 Close Lab (Choi, Close, Fenton) at
 the University of California, Riverside. Plasmid DNA
 preparations and DNA sequencing were performed in the OD
 Anderson lab (all other authors)."

FEATURES

source

BASE COUNT

141 a 182 c 156 g 120 t

ORIGIN

Query Match 5.8%; Score 47; DB 155; Length 599;
 Best Local Similarity 51.2%; Pred. No. 0.0018;
 Matches 110; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY	583	ggctccgcgtgaggaaggtgactgagacattgacgtaacggtgactgacacgac	642
DB	83	GGCCTCACCATTGTACAGGAGGAGCGGAGCGGAGCGGCGCCGCCCAAC	142
QY	643	aagaacatctactcagagatgcatgtgaccccgccctcagcgaaggtcagaa	702
DB	143	GCCACAGCCTTTCAGCGCGGCGGCTTCATGCTCGACACCAACACCCGGAG	202
QY	703	agcttcaacgagcaggttcgagagatgacgaacctgctcgaattcgagctcgac	762
DB	203	GGCAGCGGCGCATCACCAAGACTCGCGGCGGAGAGCGCTCTCCCGTCAAG	262
QY	763	aagcttgagccgacactgagacacacacacga	797
DB	263	CAGATTATGAGCGCTCCAGACCAACGACGACAA	297

RESULT 2
 BE494282 529 bp mRNA EST 02-AUG-2000
 LOCUS WHE1253_H10.P1925 Secale cereale anther cDNA library Secale cereale
 DEFINITION cDNA clone WHE1253_H10.P19, mRNA sequence.
 ACCESSION BE494282
 VERSION BE494282.1 GI:9660875
 KEYWORDS EST.
 SOURCE rye.
 ORGANISM Secale cereale
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Secale.
 1 (bases 1 to 529)
 REFERENCE Anderson O.D., Butler E., Chao S., Choi D.W., Close T.J., Fenton

ORGANISM Magnaporthe grisea
 COMMENT Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
 REFERENCE 1 (bases 1 to 796)
 AUTHORS Choi, W., Fang, E., Sasinowski, M., Wing, R. and Dean, R.A.
 TITLE Expressed sequence characterization during appressorium formation in rice blast fungus, Magnaporthe grisea
 JOURNAL Unpublished (1998)
 COMMENT Contact: Dean, R.A.
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson University, Clemson, SC 29634, USA
 Tel: 864 656 5737
 Fax: 864 656 4293
 Email: rdeane@clemson.edu
 Seq primer: T3 primer (AATTAACCCCTCACTAAGG)
 High quality sequence stop: 348.
 FEATURES
 Location/Qualifiers
 1..796
 /organism="Magnaporthe grisea"
 /strain="70-15"
 /db_xref="taxon:148305"
 /clone="mgae0005ab09f"
 /clone_lib="Magnaporthe grisea Appressorium Stage cDNA Library"
 /dev_stage="Germinated conidia on appressorium-inductive surface"
 /note="Vector: pBluescriptII SK(+). Vector: Site 1: EcoRI; Site 2: XhoI. The appressorium formation-specific cDNA library was constructed from conidia germinated for 5-8 hr on an inductive surface. The library has an average insert size of 1.5 kbp."
 BASE COUNT 175 a 268 c 198 g 146 t 9 others
 ORIGIN
 Query Match 4.7%; Score 37.8; DB 15; Length 796;
 Best Local Similarity 49.7%; Pred. No. 1.3;
 Matches 96; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
 QY 185 tcggcaagaatccggcaggttccagtcacacatcattaccggttaagccggcgacaaa 244
 |||||
 166 TCGGAAGACGACAGATCGTGTCCAGAGCAACTTCTCCAAAGGTAATGACACCGTCT 225
 |||||
 QY 245 agactagcgaagaacaccatgctgttagcccgccgcgcgataagcttccacacctag 304
 |||||
 Db 226 ATGGCCCTGGCCAGTTCTCAGACCTGCTGCAACGACGATGACACCTCCTGACTTACA 285
 |||||
 QY 305 gtctcgaatgactcgaattacgctcgcctgagctgtgacggtcaggaagtcgcgaaga 364
 |||||
 Db 286 CTCTTGACTGAGACCAAGACGACGCTCAGTGTGATGTCAACGGCAGGTGTCGCCACCC 345
 |||||
 QY 365 cggagagtgcca 377
 |||||
 Db 346 TGAAGCGCGCGCA 358
 |||||
 RESULT 5
 LOCUS A1069170 849 bp mRNA EST 09-DEC-1999
 DEFINITION mgae0005dc07f Magnaporthe grisea Appressorium Stage cDNA Library
 MAGNAPORTHE GRISEA cDNA clone mgae0005dc07f 5', mRNA sequence.
 ACCESSION A1069170
 VERSION A1069170.1 GI:3392145
 KEYWORDS EST.
 SOURCE Magnaporthe grisea.
 ORGANISM Magnaporthe grisea.
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
 REFERENCE 1 (bases 1 to 849)
 AUTHORS Choi, W., Fang, E., Sasinowski, M., Wing, R. and Dean, R.A.
 TITLE Expressed sequence characterization during appressorium formation in rice blast fungus, Magnaporthe grisea

JOURNAL Unpublished (1998)
 COMMENT Contact: Dean, R.A.
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson University, Clemson, SC 29634, USA
 Tel: 864 656 5737
 Fax: 864 656 4293
 Email: rdeane@clemson.edu
 Seq primer: T3 primer (AATTAACCCCTCACTAAGG)
 High quality sequence stop: 324.
 FEATURES
 Location/Qualifiers
 1..849
 /organism="Magnaporthe grisea"
 /strain="70-15"
 /db_xref="taxon:148305"
 /clone="mgae0005dc07f"
 /clone_lib="Magnaporthe grisea Appressorium Stage cDNA Library"
 /dev_stage="Germinated conidia on appressorium-inductive surface"
 /note="Vector: pBluescriptII SK(+). Vector: Site 1: EcoRI; Site 2: XhoI. The appressorium formation-specific cDNA library was constructed from conidia germinated for 5-8 hr on an inductive surface. The library has an average insert size of 1.5 kbp."
 BASE COUNT 203 a 290 c 198 g 152 t 6 others
 ORIGIN
 Query Match 4.7%; Score 37.8; DB 15; Length 849;
 Best Local Similarity 50.8%; Pred. No. 1.4;
 Matches 90; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
 QY 185 tcggcaagaatccggcaggttccagtcacacatcattaccggttaagccggcgacaaa 244
 |||||
 Db 196 TCGGAAGACGACAGATCGTGTCCAGAGCAACTTCTCCAAAGGTAATGACACCGTCT 255
 |||||
 QY 245 agactagcgaagaacaccatgctgttagcccgccgcgcgataagcttccacacctag 304
 |||||
 Db 256 ATGGCCCTGGCCAGTTCTCAGACCTGCTGCAACGACGATGACACCTCCTGACTTACA 315
 |||||
 QY 305 gtctcgaatgactcgaattacgctcgcctgagctgtgacggtcaggaagtcgcgaaga 361
 |||||
 Db 316 CTCTTGACTGAGACCAAGACGACGCTCAGTGTGATGTCAACGGCAGGTGTCGCCGCA 372
 |||||
 RESULT 6
 LOCUS BG103002 561 bp mRNA EST 30-JAN-2001
 DEFINITION RH122.35_H06_b1_A003 Rhizome2 (RH122) Sorghum prolinguunum cDNA, mRNA
 sequence.
 ACCESSION BG103002
 VERSION BG103002.1 GI:12617835
 KEYWORDS EST.
 SOURCE Sorghum prolinguunum.
 ORGANISM Sorghum prolinguunum.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 561)
 Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt, L.H.
 TITLE An EST database from Sorghum: Sorghum prolinguunum rhizomes
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Seq primer: JEN REV
 High quality sequence stop: 550

FEATURES

POLYA-No.

Location/Qualifiers

1..561

/organism="Sorghum propinquum"

/db_xref="taxon:132711"

/clone_lib="Rhizome2 (RHIZ2)"

/note="Organ: Rhizomes; Vector: pBluescript II from Lambda

zap II; Site_1: XhoI; Site_2: EcoRI; The library was made

from poly-A RNA in the cloning vector Lambda zap II;

clones to be sequenced were prepared by mass excision."

BASE COUNT 101 a 180 c 196 g 84 t

ORIGIN

Query Match 4.6%; Score 37.2; DB 173; Length 561;

Best Local Similarity 49.0%; Pred. No. 1.8;

Matches 99; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 161 gggtagaagtgagatattgaagtcctcgcaagaatccggcagttccagtcacaatca 220

Db 8 GGCACAGATCGACTTCGAGTCTCCGCAACGTCCTCCGCGAGCCCTACACGCTGCACA 67

Qy 221 ttacggttaagcgcgcaacaaagactagcgaaagacacatgtctgttagcccgccg 280

Db 68 CCACGCTGTCACCG 127

Qy 281 ccgataagcttcacacactacgctcgaatgactccgaatcagtcgctgagctg 340

Db 128 CCACGACGACTTCACACCTACTCCGTGTGTGGAACCGCGACGATGATCTTGCCGC 187

Qy 341 ttgacggtcagaagatccgca 362

Db 188 TGGACGCGACCCCGATCCGCGA 209

RESULT 7

AI397807

LOCUS AI397807 552 bp mRNA EST 10-JAN-2001

DEFINITION NCM9C473 Mycelial Neurospora crassa cDNA clone NM9C4 5', mRNA

ACCESSION AI397807

VERSION AI397807.1 GI:4240892

KEYWORDS EST.

SOURCE Neurospora crassa.

ORGANISM Neurospora crassa.

REFERENCE 1 (bases 1 to 552)

AUTHORS Nelson, M.A., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L.,

Leonard, P.M., Mitchell, J., Armijo, A.M., Bean, L., Blueyes, E.,

Cushling, T., Errett, A., Flenahy, M., Gorman, M., Judson, K., Miller, R.,

Ortega, J., Pavlova, I., Perea, J., Todisco, S., Trujillo, R.,

Valentine, J., Wells, A., Werner, Washburne, M., Yazzie, S. and Nativg

, D.O.

TITLE

Expressed sequences from conidial, mycelial, and sexual stages of

Neurospora crassa

JOURNAL Fungal Genet. Biol. 21, 348-363 (1997)

MEDLINE 97435549

COMMENT Contact: Nativg, D.O./Nelson, M.A.

Department of Biology

University of New Mexico

Cascabel Hall, Albuquerque, NM 87131, USA

Tel: 505 277 3411

Fax: 505 277 0304

Email: npebbiology.unm.edu.

Location/Qualifiers

1..552

/organism="Neurospora crassa"

/strain="74-OR23-IV A (FGSC 2489)"

/db_xref="taxon:5141"

/clone_lib="NM9C4"

/clone_lib="Mycelial"

/sex="Mating type A"

FEATURES

/tissue_type="Mycelium"

/dev_stage="Mycelium"

/lab_host="E. coli"

/note="Vector: pBluescript SK (-); Site1: EcoRI; Site2:

XhoI; 28 sucrose for 24 hours. cDNA directionally cloned

into pBluescript SK(-) using the Uni-ZAP XR vector system

(Stratagene, La Jolla, CA)."

BASE COUNT 127 a 137 c 172 g 116 t

ORIGIN

Query Match 4.6%; Score 37; DB 19; Length 552;

Best Local Similarity 46.9%; Pred. No. 2.1;

Matches 115; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

Qy 455 tggataatcaaacgcttcgcttcagttccagttcaacggttcaggttataatata 514

Db 293 TCGAAGACTACGTGATCCGATCGACAGCGTCACACACTCCGCGAGACTACGCGATCG 352

Qy 515 ccgagcgccagggcggaagcgcgacgacttaccgttaccgtgaccgcaatttga 574

Db 353 GGGTCGACAAAGCGTGTGCGGTGAGGACTATGTATGAGGTCGACAGCGTCACGCTG 412

Qy 575 cgttatgctccgctggtggcaagggtgactgacattgaacgttgcagc 634

Db 413 GTGTTGAGGACTACGCGCATCGGGGTGATTAAGCGTGTGTCGAGGACTACGTATCC 472

Qy 635 taaccgacagaacatctctccagagatgcatgttgcctcgccaccgcaag 694

Db 473 CGATCGACAAAGCGCACACTCCGTCGAAGACTATGTATGAAGTCGACAGCTTAAT 532

Qy 695 gtccag 699

Db 533 GACCG 537

RESULT 8

BF861648

LOCUS BF861648 747 bp mRNA EST 19-JAN-2001

DEFINITION 963025E08.Y1 C. reinhardtii CC-1690, Stress condition I, normalized

Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BF861648

VERSION BF861648.1 GI:12251785

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii.

REFERENCE 1 (bases 1 to 747)

AUTHORS Grossman, A., Davies, J., Federpiet, N., Harris, E., Hauser, C.,

Lefebvre, P., McDermott, J.P., Shreger, J., Silflow, C. and Stern, D.

Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants, project phase 3

JOURNAL Unpublished (2000)

COMMENT Contact: Charles Hauser

Duke University

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

Location/Qualifiers

1..747

/organism="Chlamydomonas reinhardtii"

/strain="CC-1690 wild type mt+ 219r"

/db_xref="taxon:3055"

/clone_lib="C. reinhardtii CC-1690, Stress condition I,

normalized, Lambda zap II"

/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:

XhoI; This library, constructed by John Davies and Jeffrey

McDermott, combines cDNAs from CC-1690 cells grown to

mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,

```
FEATURES
source
location/Qualifiers
1. .408
/organism="Homo sapiens"
/db_xref="GDB:3849557"
/db_xref="taxon:9606"
/cclone="IMAGE:221564"
```

```
Seq primer: JEN REV
High quality sequence stop: 536
POLYA-No.
Location/Qualifiers
1.604
source
/organism="Sorghum bicolor"
```


Fri Aug 24 10:05:31 2001

us-09-654-652a-5.rst

Page 11

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 15:04:24 ; Search time 2906.88 Seconds
(without alignments)
4278.151 Million cell updates/sec

Title: US-09-654-652A-5

Perfect score: 804
Sequence: 1 atggttagcgcaaggattt.....accaccaccaccaccactga 804

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pi1:*
13: gb_pi2:*
14: gb_pi3:*
15: gb_pi4:*
16: gb_ba1:*
17: gb_ba2:*
18: gb_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pi1:*
48: em_pi2:*
49: em_pi3:*
50: em_pi4:*
51: em_pi5:*
52: em_pi6:*
53: em_pi7:*
54: em_pi8:*
55: em_pi9:*
56: em_pi10:*
57: em_pi11:*
58: em_pi12:*
59: em_pi13:*
60: em_pi14:*
61: em_pi15:*
62: em_pi16:*
63: em_pi17:*
64: em_pi18:*
65: em_pi19:*
66: em_pi20:*
67: em_pi21:*
68: em_pi22:*
69: em_pi23:*
70: em_pi24:*
71: em_pi25:*
72: em_pi26:*
73: em_pi27:*
74: em_pi28:*
75: em_pi29:*
76: em_pi30:*
77: em_pi31:*
78: em_pi32:*
79: em_pi33:*
80: em_pi34:*
81: em_pi35:*
82: em_pi36:*
83: em_pi37:*
84: em_pi38:*
85: em_pi39:*
86: em_pi40:*
87: em_pi41:*
88: em_pi42:*
89: em_pi43:*
90: em_pi44:*
91: em_pi45:*
92: em_pi46:*
93: em_pi47:*
94: em_pi48:*
95: em_pi49:*
96: em_pi50:*
97: em_pi51:*
98: em_pi52:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	741	92.2	1426	3	MTBGLUC
2	64	8.0	4411	9	AR069366 Sequence
3	64	8.0	5443	9	AR069362 Sequence
4	64	8.0	5616	9	AR069364 Sequence
5	61	7.6	2735	9	AR054309 Sequence
6	55	6.8	6363	9	AR069367 Sequence
7	45	5.6	98	9	AX026763 Sequence
8	43	5.3	117	56	ASTOOL10 Artificial

9	43	5.3	408	56	ASAUTOVAC	Y09276 Artificial
10	43	5.3	5549	56	AF0977413	Y09275 Artificial
11	42	5.2	306	56	ASTCELL90	Y09275 Artificial
12	42	5.2	357	9	AX025780	AX025780 Sequence
13	42	5.2	586	9	AX025776	AX025776 Sequence
14	42	5.2	6619	9	AX028564	AX028564 Sequence
15	36	4.5	1815	9	A79318	A79318 Sequence 6
16	36	4.5	1815	9	AX074521	AX074521 Sequence
17	36	4.5	5639	56	AF117644	AF117644 Expressio
18	36	4.5	5631	56	AF117645	AF117645 Expressio
19	36	4.5	5727	56	AF116269	AF116269 Expressio
20	34	4.2	948	9	AR054722	AR054722 Sequence
21	34	4.2	948	9	AR061718	AR061718 Sequence
22	34	4.2	948	9	AR061959	AR061959 Sequence
23	34	4.2	948	9	AR063660	AR063660 Sequence
24	34	4.2	948	9	AR086141	AR086141 Sequence
25	34	4.2	948	9	AR089566	AR089566 Sequence
26	34	4.2	948	9	AR093656	AR093656 Sequence
27	34	4.2	948	10	I38603	I38603 Sequence 31
28	34	4.2	963	9	AR062083	AR062083 Sequence
29	34	4.2	273	9	AR089600	AR089600 Sequence
30	32	3.6	273	56	ASTOOL40	Y09273 Artificial
31	28	3.5	738	10	AX073930	AX073930 Sequence
32	28	3.5	981	9	AR099878	AR099878 Sequence
33	28	3.5	3259	56	XX016281	U16281 Cloning vec
34	28	3.5	4779	56	XX061572	U16722 Cloning vec
35	28	3.5	6353	9	AX063157	AX063157 Sequence
36	27	3.4	294	9	AR099877	AR099877 Sequence
37	27	3.4	294	9	AR099879	AR099879 Sequence
38	27	3.4	3027	88	AF007748	AF007748 Homo sapi
39	27	3.4	4946	3	PPVIRE5	Y07702 Plasmid pPV
40	27	3.4	5502	3	AX063158	AX063158 Sequence
41	26	3.2	780	56	SYNN198SCF	L43067 Synthetic s
42	26	3.2	5319	56	AF161077	AF161077 Cloning v
43	26	3.2	6372	56	AF117643	AF117643 Expressio
44	26	3.2	6567	56	AF053733	AF053733 Expressio
45	26	3.2	9915	56	AF055298	AF055298 Expressio

ALIGNMENTS

	/codon_start=1
	/transl_table=1
	/protein_id="AA24896.1"
	/db_xref="GI:148576"
	/translation="MIKIKTAVKSALVAANAALITNVSAKDFSGAELTYLLEVOYG KEEARMKMAASCPVSSMPFLYONSESLADRPWEVDIELVGNPGSFOSNIITGAG AQKTSEKHHAIVSPADAQFHTYGLERTPVVRMTYDQAEVRKTEGGVSNLTGTGGR FLNTSSESAAWVGOFDESKLPLEFQILMVVYRTTPGQEGGSQFTLDITDNEDFDG SRWKGDWTFDGENRVDLTDKNYISRDMLIALTRKQESFNQGVPPDDPAQSSES APASSSSVPASSSSVPASSSSSAEPFPSSSATNALHGMRTTPVAKCHKRRLLVNAAKAK VNPGRHKRYRVNFEEH" 145 225
sig_peptide	/note="1,3-1,4-beta-D-glucan 4-glucanohydrolase signal peptide 226..1191 /note="1,3-1,4-beta-D-glucan 4-glucanohydrolase"
mat_peptide	226..1191
BASE COUNT	371 a 346 c 335 g 374 t
ORIGIN	
Query Match	92.2%; Score 741; DB 3; Length 1426;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 741; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Oy	4 gttacgcgaagatttagcgglytccgaactctaacgcttagaagaagtccagtaagyt 63
Dd	217 GTTAGCGCAAGAAGATTATTACGGCTCCGAACCTACGTTAGAAGAAGTTCACTACGCT 276
Oy	64 aagtttgaaaccgcatataagatgtccgacgccgatccggaaacagtcacgttccttc 123
Dd	277 AAGTTTGAAGCCCGCATATAAAGATGGACCACCCATCGGGGAACAAGTCAGTTCATGTCCTC 336
Oy	124 taccaaatgtgttcggaatccgcgaatgtgaaggcccctgtgtagaagtgtgaatgaagt 183
Dd	337 TACCGAATGTGTTCCGAATAATGCCCATGGAAGCCCTGCGTAGAAGTGCATATTGAAGTT 396
Oy	184 ctccgcaagaatccggcgagittccagtcacaacatcattaccggtlaaggccgcgcgcaaa 243

```

RESULT 1
FIBRLUC
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

FIBRLUC 1426 bp DNA BCT 26-APR-1993
F.succinogenes 1,3,1,4-beta-D-glucan 4-glucanohydrolase gene,
complete cds.
M33676 M3311
M33676.1 GI:148575
1,4-beta-D-glucan 4-glucanohydrolase; missed-linkage beta-glucanase
F.succinogenes (strain S85) DNA, clone FJ15.
Fibrobacter succinogenes
Bacteria; Fibrobacter/Acidobacteria group; Fibrobacter group;
Fibrobacter.
1 (bases 1 to 1426)
Teather,R.M. and Ertle,J.D.
DNA sequence of a Fibrobacter succinogenes mixed-linkage
beta-glucanase (1,3,1,4-beta-D-glucan 4-glucanohydrolase) gene
J.Bacteriol. 172, 3837-3841 (1990)
90299607
Draft entry and computer-readable sequence for [1] kindly submitted
by R.M.Teather, 11-APR-1990.
Location/Qualifiers
1. 1426
/organism="Fibrobacter succinogenes"
/db_xref="taxon:83"
62. .66
85. .90
132. .137
/note="ribosome binding site"
145. 1194
/note="1,3,1,4-beta-D-glucan 4-glucanohydrolase precursor
(EC 3.2.1.73)"

```

			/codon_start=1	
			/transl_table=11	
			/protein_id="AA024896.1"	
			/db_xref="gi:148576"	
			/translation="MNIKTPTAVKSAALVAIAAAAAAALTNNVSAKDFSGAEIYLTLEEVQGIKEERMKMAASSTGVSSFMETLXONGSEFNPADGPPWEVDIVELGRKPGEFSQNIITGKGAOKTSKKHHVAPSPADAOFHYGLKEMTPNVRYRMVDQEARTRKGCGVSLTGTCIGRFLNLSSESANAYGQPFDESKLPDFOPIIMVKYKYKTTPQGGEGSDFTLDWINDPDTFGSRWKGDWTFEDGNRVDLTDKNITYRDMLLIATLRKQESFNQGVPPDDPAPOSSSAPASSSVPSAPSSVPASSSSAFVFPSSSTYNIAIHQMRTTPAVAKEHRLVNAGKARKVNPNGHRYRNVEFH"	
			VNPNGHRYRNVEFH"	
			145..225	
	sig_peptide		/note="1,3-1,4-beta-D-glucan 4-glucanohydrolase signal peptide"	
	mat_peptide		226..1191	
			/note="1,3-1,4-beta-D-glucan 4-glucanohydrolase"	
BASE COUNT	371 a	346 c	335 g	374 t
ORIGIN				
Query Match	92.2%	Score 741;	DB 3;	Length 1426;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 741;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Oy	4	gttaacgcgaagaatttagcggctccgaactctacaqttagaagaagtcagtaagct	63	
Dd	217	gTTTAACGCgAAAAGATTtTAGCGGTCCCAACTCTAACGTTAgAGAAGATTCACTACGCT	276	
Oy	64	aagcttgaaaccgcatatgaagtgcgaacgcgatccggaacagtcagtcagcttcctc	123	
Dd	277	AGGTTTGAAGCCCCGATtTAGAGATGGCACCCGATGGGAAACAGTCAGTTCATGTTCTC	336	
Oy	124	taccgaatggtctcgaaatcgcgatggaagccctggtagaaglygatatgtgaatt	183	
Dd	337	TACAGAAATGGCTTCGGAATCCCGCATGGAGGCCCTGGTAgAAGTGAATTTGAAGTT	396	
Oy	184	ctcggcaagaataccggaggatlttcagfccaacatcatctaccgttaaggcggcgacaa	243	
Dd	397	cTCCGCAAGAAATCCGGGAGTtTCCAAGTCAACAATCAATACCAGGTAAGGCGGGCACAA	456	
Oy	244	aagactagcgaaaagcacacatgctgtlttagcccgccgcgcga tcaggtcttcacacctac	303	
Dd	457	AAGACTTAGCGAAACACACCATGCTTTAGCCCGCGCGCATCAAGCTTTCACACCTAC	516	
Oy	304	ggctcgaatcgactcgaatlaogtccgctfgnactglttgaagylcaagaaagtcgcgaag	363	
Dd	517	GGTCTCGAATGACTCCGAATTAAGTCCGCTGGAGACTGTTCAGGGTCAAGGAAGTCCGCAAG	576	
Oy	364	acggaagtlgscgaagtttcaacttgcacgaigtacaagggaccgctttaaccttgg	423	
Dd	577	ACGGAAAGTGGCCAAAGTtTCCAACTTGACAGGTACACGGGACACTCTCGTTTAACCTTTGG	636	
Oy	424	tcgcttggaagtgccgcttgggttbgccagcttcgaatgaatcaaaagcttcgccttccag	483	
Dd	637	TCGTCTGAGAGTGGCGGCTTGGGTTGGCCAGTTCATGATCAAAGCTTCCGCTTTCCAG	696	
Oy	484	ttcatcaactggttcaaggttatataagatacgcgcggccgaagggcgaaagcggcagcgac	543	
Dd	697	TTTCATCAACTGGGTCAGAGTtTATTAAGATATACGCCGGGCGCAAGGGGCAAGCGACAC	756	
Oy	544	tttaagcttgaactggaaccgacaatttgaacagcttgtatggtctccgctgggcaaggt	603	
Dd	757	TTTTAGCTTTGACTGGACCGCAATTTTGCACAGTtTGTGCTCTCCCGTGGGGCAAGGGT	816	
Oy	604	gacitggaacttgaacggaacccgttgcgaacctaacgcgaagaacatctactcaagat	663	
Dd	817	GACTGGAACTTTGAGAGGTACCGTGTGCACTTCACCGCAAGAACATCTTACTCCAAGAT	876	
Oy	664	ggacgttgcattccttcgccctcacccgcaaaggttaagaaagcttcaacygcaggtctcg	723	
Dd	877	GGCAAGTGTGATCTCTGCCCTCACCCGCAAGGTCAAGAAACCTTCACAGCGCAGTTCG	936	
Oy	724	agagatlgacgaacctgctcg	744	

Db 937 AGAGATGACGAACTGCTCGG 957

RESULT 2
LOCUS AR069366/c 4411 bp DNA PAT 18-FEB-2000
DEFINITION Sequence 5 from patent US 5891637.
ACCESSION AR069366
VERSION AR069366.1 GI:7220254
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 4411)
AUTHORS Ruppert,S.J.W.
TITLE Construction of full length cDNA libraries
JOURNAL Patent: US 5891637-A 5 06-APR-1999;
FEATURES
Source 1..4411
Location/Qualifiers
BASE COUNT 1065 a 1156 c 1124 g 1066 t
ORIGIN

Query Match 8.0%; Score 64; DB 9; Length 4411;
Best Local Similarity 100.0%; Pred. No. 3.1e-22;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 tcggaattcgagctcgctgcgacaagcttgcgcccgcacatcgagacaccacacacca 800
|||||
Db 1715 TCGGAATTGCGAGCTCGCTGCACAGCTTGCGCGCGCAGCTGAGACACCAACCA 1656

QY 801 ctga 804
|||||
Db 1655 CTGA 1652

RESULT 3
LOCUS AR069362 5443 bp DNA PAT 18-FEB-2000
DEFINITION Sequence 1 from patent US 5891637.
ACCESSION AR069362
VERSION AR069362.1 GI:7220250
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 5443)
AUTHORS Ruppert,S.J.W.
TITLE Construction of full length cDNA libraries
JOURNAL Patent: US 5891637-A 1 06-APR-1999;
FEATURES
Source 1..5443
Location/Qualifiers
BASE COUNT 1268 a 1441 c 1446 g 1288 t
ORIGIN

Query Match 8.0%; Score 64; DB 9; Length 5443;
Best Local Similarity 100.0%; Pred. No. 3e-22;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 tcggaattcgagctcgctgcgacaagcttgcgcccgcacatcgagacaccacacacca 800
|||||
Db 5244 TCGGAATTGCGAGCTCGCTGCACAGCTTGCGCGCGCAGCTGAGACACCAACCA 5303

QY 801 ctga 804
|||||
Db 5304 CTGA 5307

RESULT 4

AR069364
LOCUS AR069364 5616 bp DNA PAT 18-FEB-2000
DEFINITION Sequence 3 from patent US 5891637.
ACCESSION AR069364
VERSION AR069364.1 GI:7220252
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 5616)
AUTHORS Ruppert,S.J.W.
TITLE Construction of full length cDNA libraries
JOURNAL Patent: US 5891637-A 3 06-APR-1999;
FEATURES
Source 1..5616
Location/Qualifiers
BASE COUNT 1306 a 1478 c 1492 g 1340 t
ORIGIN

Query Match 8.0%; Score 64; DB 9; Length 5616;
Best Local Similarity 100.0%; Pred. No. 3e-22;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 tcggaattcgagctcgctgcgacaagcttgcgcccgcacatcgagacaccacacacca 800
|||||
Db 5417 TCGGAATTGCGAGCTCGCTGCACAGCTTGCGCGCGCAGCTGAGACACCAACCA 5476

QY 801 ctga 804
|||||
Db 5477 CTGA 5480

RESULT 5
LOCUS AR054309 2735 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5837240.
ACCESSION AR054309
VERSION AR054309.1 GI:5979886
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2735)
AUTHORS Lee,C.K., Monath,T.P., Ackerman,S.K., Thomas,W.D., Soman,G., Kleanthous,H., Weitzin,R.A., Pappo,J., Ermak,T., Guirakhoo,F., Bhagat,H. and Sussman,I.
TITLE Multimeric, recombinant urease vaccine
JOURNAL Patent: US 5837240-A 1 17-NOV-1998;
FEATURES
Source 1..2735
Location/Qualifiers
BASE COUNT 875 a 562 c 635 g 663 t
ORIGIN

Query Match 7.6%; Score 61; DB 9; Length 2735;
Best Local Similarity 100.0%; Pred. No. 1.3e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 744 gaattcgagctcgctgcgacaagcttgcgcccgcacatcgagacaccacacacactg 803
|||||
Db 2564 GAATTGAGCTCGCTGCACAGCTTGCGCGCGCAGCTGAGACACCAACCACTG 2623

QY 804 a 804
|
Db 2624 A 2624

RESULT 6
LOCUS AR069367 6363 bp DNA PAT 18-FEB-2000
DEFINITION Sequence 6 from patent US 5891637.

FEATURES	(IT)
source	location/Qualifiers 1. .586 /organization="SEN virus" /db_xref="taxon:136966"
BASE COUNT	155 a 172 c 139 g 119 t
ORIGIN	1 others

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 15:01:15 ; Search time 157.23 Seconds
(without alignments)
3210.792 Million cell updates/sec

Title: US-09-654-652A-5

Perfect score: 804
Sequence: 1 atggttagcgcaagaattt.....accaccaccaccactga 804

Scoring table:
OLIGO_NUC
Gapox 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq_0601:*

1: /cgnl_9/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /cgnl_9/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /cgnl_9/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /cgnl_9/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /cgnl_9/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /cgnl_9/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /cgnl_9/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /cgnl_9/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /cgnl_9/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /cgnl_9/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /cgnl_9/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /cgnl_9/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /cgnl_9/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /cgnl_9/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /cgnl_9/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /cgnl_9/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /cgnl_9/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /cgnl_9/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /cgnl_9/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /cgnl_9/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /cgnl_9/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /cgnl_9/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	8.0	4411	20	AAx28088
2	64	8.0	5616	20	AAx28085
3	61	7.6	2007	20	AAx80369
4	61	7.6	2007	20	AAx80370
5	61	7.6	2007	20	AAx80366
6	61	7.6	2007	20	AAx80367
7	61	7.6	2007	20	AAx80368
8	61	7.6	2735	17	AAx44351
9	61	7.6	5248	16	AAx08489
10	61	7.6	5248	17	AAx41898
11	55	6.8	6363	20	AAx28091

12	50	6.2	59	22	AAx74635
13	42	5.2	357	21	AAx61167
14	42	5.2	586	21	AAx61166
15	42	5.2	1186	20	AAx29726
16	42	5.2	1247	20	AAx29725
17	42	5.2	6619	20	AAx63789
18	41	5.1	423	20	AAx19396
19	41	5.1	423	20	AAx19184
20	39	4.9	550	21	AAx71987
21	37	4.6	607	21	AAx64807
22	36	4.5	241	21	AAx94064
23	36	4.5	1815	17	AAx39341
24	36	4.5	1815	19	AAx15138
25	34	4.2	948	16	AAx080748
26	34	4.2	948	17	AAx27687
27	34	4.2	948	18	AAx70349
28	34	4.2	948	18	AAx76649
29	34	4.2	948	19	AAx65799
30	34	4.2	948	19	AAx53873
31	34	4.2	948	20	AAx63425
32	34	4.2	963	18	AAx76626
33	34	4.2	963	19	AAx65831
34	34	4.2	963	19	AAx53941
35	29	3.6	783	16	AAx08490
36	28	3.5	738	22	AAx25015
37	28	3.5	981	21	AAx51210
38	28	3.5	6353	22	AAx68848
39	28	3.5	7375	21	AAx09770
40	28	3.5	7375	21	AAx98759
41	27	3.4	294	21	AAx51209
42	27	3.4	294	21	AAx51211
43	27	3.4	980	17	AAx42665
44	27	3.4	3572	20	AAx19454
45	27	3.4	3572	20	AAx19242

ALIGNMENTS

RESULT 1	
AAx28088/c	
ID	AAx28088 standard; DNA; 4411 BP.
AC	AAx28088;
XX	
XX	10-JUN-1999 (first entry)
DE	Plasmid PACYC.tac.
XX	
KW	Mult reverse transcriptase; cDNA production; cDNA library production;
XX	ss.
OS	Synthetic.
XX	
PN	US5891637-A.
XX	
PD	06-APR-1999.
XX	
PF	15-SEP-1997; 97US-0929967.
XX	
PR	15-OCT-1996; 96US-0732861.
PR	15-SEP-1997; 97US-0929967.
XX	
PA	(GENTH) GENENTECH INC.
XX	
PI	Ruppert SJ;
XX	
DR	WPI; 1999-253852/21.
XX	
PT	Producing double stranded cDNA molecules from mRNA transcripts and a
PT	novel xenotropic murine leukemia virus reverse transcriptase -
XX	useful for constructing cDNA libraries

PS Example 5; Fig 9; 70pp; English.
XX
CC This sequence represents a plasmid used to test the method of
CC the invention. The method is for the production of double
CC stranded cDNA (dsCDNA) molecules from mRNA transcripts and a xenotropic
CC murine leukemia virus reverse transcriptase (xm-MuLV). The method, the
CC host cells, reverse transcriptases and nucleic acids disclosed may be
CC used to produce high quality, normalized, full-length, directionally
CC cloned, cDNA libraries. Use of the method to construct cDNA libraries
CC ensures that the reverse transcriptase is able to extend the first strand of
CC the cDNA up to the ultimate 5' end of the mRNA, therefore ensuring that
CC all sequences are represented in the library, whereas previously, some
CC sequences would be lost due to incomplete transcription (e.g. as a result
CC of hairpin loop formation) or mRNA degradation (e.g. as a result of RNase
CC H activity). Also, carrying the process out in cells means that enzymes
CC and proteins present in the cell will repair any mistakes or nicks in the
CC product DNA. Modifying the mRNA for insertion into a vector by adding the
CC 5' cap provides a convenient way of priming it for cDNA synthesis. Also,
CC the method provides a fast and clean way of synthesizing cDNA
CC clones.
XX
SQ Sequence 4411 BP; 1065 A; 1156 C; 1124 G; 1066 T; 0 other;

Query Match 8.0%; Score 64; DB 20; Length 4411;
Best Local Similarity 100.0%; Pred. No. 9.2e-22;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 741 tcggaattcgagctcgctgcacaagcttgcgcccagctcgagacaccaccaccca 800
|||||
1715 TCGGAATTCGAGCTCGCTGCACAGCTTGGCGCGCAGCTGAGACACACACACCA 1656

QY 801 ctga 804
|||||
DB 1655 CTGA 1652

RESULT 2
AAAX28085
ID AAAX28085 standard; DNA; 5616 BP.
XX
AC AAAX28085;

XX 10-JUN-1999 (first entry)

DE Plasmid pET21.lac.

XX MULV reverse transcriptase; dsCDNA production; cDNA library production;
KM ss.

OS Synthetic.

PN US5891637-A.

PD 06-APR-1999.

PF 15-SEP-1997; 97US-0929967.

PR 15-OCT-1996; 96US-0732861.

PR 15-SEP-1997; 97US-0929967.

PA (GENTH) GENENTECH INC.

PI Ruppert SJ;

DR WPI; 1999-253852/21.

XX Producing double stranded cDNA molecules from mRNA transcripts and a
XX novel xenotropic murine leukemia virus reverse transcriptase -
XX useful for constructing cDNA libraries
PS Example 5; Fig 7; 70pp; English.
XX

CC This sequence represents a plasmid used to test the method of
CC the invention. The method is for the production of double
CC stranded cDNA (dsCDNA) molecules from mRNA transcripts and a xenotropic
CC murine leukemia virus reverse transcriptase (xm-MuLV). The method, the
CC host cells, reverse transcriptases and nucleic acids disclosed may be
CC used to produce high quality, normalized, full-length, directionally
CC cloned, cDNA libraries. Use of the method to construct cDNA libraries
CC ensures that the reverse transcriptase is able to extend the first strand of
CC the cDNA up to the ultimate 5' end of the mRNA, therefore ensuring that
CC all sequences are represented in the library, whereas previously, some
CC sequences would be lost due to incomplete transcription (e.g. as a result
CC of hairpin loop formation) or mRNA degradation (e.g. as a result of RNase
CC H activity). Also, carrying the process out in cells means that enzymes
CC and proteins present in the cell will repair any mistakes or nicks in the
CC product DNA. Modifying the mRNA for insertion into a vector by adding the
CC 5' cap provides a convenient way of priming it for cDNA synthesis. Also,
CC the method provides a fast and clean way of synthesizing cDNA
CC clones.
XX
SQ Sequence 5616 BP; 1306 A; 1478 C; 1492 G; 1340 T; 0 other;

Query Match 8.0%; Score 64; DB 20; Length 5616;
Best Local Similarity 100.0%; Pred. No. 9e-22;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 tcggaattcgagctcgctgcacaagcttgcgcccagctcgagacaccaccaccca 800
|||||
DB 5417 TCGGAATTCGAGCTCGCTGCACAGCTTGGCGCGCAGCTGAGACACACACACCA 5476

QY 801 ctga 804
|||||
DB 5477 CTGA 5480

RESULT 3
AAAX80369
ID AAAX80369 standard; cDNA; 2007 BP.
XX
AC AAAX80369;

XX 07-SEP-1999 (first entry)

DE HCV NS4A-NS3 complex encoding cDNA SEQ ID:NO:119.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;

KM NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;

XX hydrophobic domain; covalent complex; detection; inhibitor; ss.

OS Hepatitis C virus.

PN WO9928482-A2.

PD 10-JUN-1999.

PF 24-NOV-1998; 98WO-US24528.

PR 28-JUL-1998; 98US-0094331.

PR 28-NOV-1997; 97US-0067315.

PA (SCHE) SCHERING CORP.

PI Malcolm BA, Taremi SS, Weber PC, Yao N;

DR WPI; 1999-385385/32.

XX New hepatitis C virus covalent complexes
XX Disclosure; Page 204-207; 211pp; English.
PS
XX
XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV

CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.

SO Sequence 2007 BP; 413 A; 621 C; 563 G; 410 T; 0 other;

Query Match 7.6%; Score 61; DB 20; Length 2007;
Best Local Similarity 100.0%; Pred. No. 2.8e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 gaattgagctccgtgcagaagcttgccgcgcactgcagaccaccacaccacactg 803
|
DB 1942 gaattgagctccgtgcagaagcttgccgcgcactgcagaccaccacaccacactg 2001

OY 804 a 804
|
DB 2002 a 2002

RESULT 4

AAx80370
ID AAX80370 standard; cDNA; 2007 BP.

AC AAX80370;

DT 07-SEP-1999 (first entry)

DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:120.

DE HCV; hepatitis C virus; single chain recombinant complex; linker;

KM NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;

KM hydrophobic domain; covalent complex; detection; inhibitor; ss.

OS Hepatitis C virus.

OS Synthetic.

PN WO928482-A2.

PD 10-JUN-1999.

PF 24-NOV-1998; 98WO-US24528.

PR 28-JUL-1998; 98US-0094331.

PR 28-NOV-1997; 97US-0067315.

PA (SCHE) SCHERING CORP.

PI Malcolm BA, Taremi SS, Weber PC, Yao N;

PI WPI; 1999-385385/32.

PT New hepatitis C virus covalent complexes

PS Disclosure; Page 207-210; 211pp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent

CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.

SO Sequence 2007 BP; 413 A; 622 C; 563 G; 409 T; 0 other;

Query Match 7.6%; Score 61; DB 20; Length 2007;
Best Local Similarity 100.0%; Pred. No. 2.8e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 gaattgagctccgtgcagaagcttgccgcgcactgcagaccaccacaccacactg 803
|
DB 1942 gaattgagctccgtgcagaagcttgccgcgcactgcagaccaccacaccacactg 2001

OY 804 a 804
|
DB 2002 a 2002

RESULT 5

AAx80366
ID AAX80366 standard; cDNA; 2007 BP.

AC AAX80366;

DT 07-SEP-1999 (first entry)

DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:116.

DE HCV; hepatitis C virus; single chain recombinant complex; linker;

KM NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;

KM hydrophobic domain; covalent complex; detection; inhibitor; ss.

OS Hepatitis C virus.

OS Synthetic.

PN WO928482-A2.

PD 10-JUN-1999.

PF 24-NOV-1998; 98WO-US24528.

PR 28-JUL-1998; 98US-0094331.

PR 28-NOV-1997; 97US-0067315.

PA (SCHE) SCHERING CORP.

PI Malcolm BA, Taremi SS, Weber PC, Yao N;

PI WPI; 1999-385385/32.

PT New hepatitis C virus covalent complexes

PS Disclosure; Page 194-197; 211pp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.

SO Sequence 2007 BP; 415 A; 619 C; 564 G; 409 T; 0 other;

Query Match 7.6%; Score 61; DB 20; Length 2007;
Best Local Similarity 100.0%; Pred. No. 2.8e-20;

Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 gaattcgagctcgctgcagaaagcttgcggcgacactgcagaccaccaccaccactg 803

Db 1942 gaattcgagctcgctgcagaaagcttgcggcgacactgcagaccaccaccaccactg 2001

OY 804 a 804

Db 2002 a 2002

RESULT 6

AAx80367

ID AAX80367 standard; cDNA; 2007 BP.

XX AAX80367;

DT 07-SEP-1999 (first entry)

DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:117.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;

KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;

KM hydrophobic domain; covalent complex; detection; inhibitor; ss.

XX Hepatitis C virus.

OS Synthetic.

XX WO9928482-A2.

XX 10-JUN-1999.

PD 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

PR 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.

PI Malcolm BA, Taremi SS, Weber PC, Yao N;

DR WPI: 1999-385385/32.

PT New hepatitis C virus covalent complexes

XX Disclosure; Page 197-200; 21pp; English.

CC The present invention describes a covalent hepatitis C virus (HCV)

CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV

CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the

CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker

CC to the amino terminus of the HCV NS3 protease domain. The present

CC sequence encodes an example of the above complex. The covalent

CC NS4A-NS3 complexes are useful for structural determination and

CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.

CC They can also be used for detecting inhibitors of the protease activity,

CC the helicase activity and the ATPase activity of NS3. The covalent

CC NS4A-NS3 complexes are more soluble, stable and active than the non-

CC covalent protease-peptide complexes previously available.

XX Sequence 2007 BP; 415 A; 619 C; 564 G; 409 T; 0 other;

Query Match 7.6%; Score 61; DB 20; Length 2007;

Best Local Similarity 100.0%; Pred. No. 2.8e-20;

Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 gaattcgagctcgctgcagaaagcttgcggcgacactgcagaccaccaccaccactg 803

Db 1942 gaattcgagctcgctgcagaaagcttgcggcgacactgcagaccaccaccaccactg 2001

OY 804 a 804

Db 2002 a 2002

RESULT 7

ID AAX80368 standard; cDNA; 2007 BP.

XX AAX80368;

DT 07-SEP-1999 (first entry)

DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:118.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;

KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;

KM hydrophobic domain; covalent complex; detection; inhibitor; ss.

XX Hepatitis C virus.

OS Synthetic.

XX WO9928482-A2.

XX 10-JUN-1999.

PD 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

PR 28-NOV-1997; 97US-0067315.

XX (SCHE) SCHERING CORP.

PI Malcolm BA, Taremi SS, Weber PC, Yao N;

DR WPI: 1999-385385/32.

PT New hepatitis C virus covalent complexes

XX Disclosure; Page 201-203; 21pp; English.

CC The present invention describes a covalent hepatitis C virus (HCV)

CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV

CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the

CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker

CC to the amino terminus of the HCV NS3 protease domain. The present

CC sequence encodes an example of the above complex. The covalent

CC NS4A-NS3 complexes are useful for structural determination and

CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.

CC They can also be used for detecting inhibitors of the protease activity,

CC the helicase activity and the ATPase activity of NS3. The covalent

CC NS4A-NS3 complexes are more soluble, stable and active than the non-

CC covalent protease-peptide complexes previously available.

XX Sequence 2007 BP; 414 A; 620 C; 563 G; 410 T; 0 other;

Query Match 7.6%; Score 61; DB 20; Length 2007;

Best Local Similarity 100.0%; Pred. No. 2.8e-20;

Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 gaattcgagctcgctgcagaaagcttgcggcgacactgcagaccaccaccaccactg 803

Db 1942 gaattcgagctcgctgcagaaagcttgcggcgacactgcagaccaccaccaccactg 2001

OY 804 a 804

Db 2002 a 2002

RESULT 8

ID AAT44351 standard; cDNA; 2735 BP.

XX AAT44351;

AC AAT44351;

XX 11-FEB-1997 (first entry)
 DT
 XX
 DE H. pylori ureA + ureB gene locus in PORV214.
 XX
 KW urease; urea gene; ureB gene; vaccine; ds.
 XX
 OS Chimeric Helicobacter strain CPM630;
 OS Chimeric bacteriophage T7.
 XX
 FH
 FT Key
 FT Location/Qualifiers
 FT 1..16
 FT /tag= a
 FT /note= "T7 promoter provides transcription
 FT initiation for the urease genes"
 FT misc_signal
 FT 33..43
 FT /tag= b
 FT /function= operator
 FT /note= "Lac operator provides inducible expression
 FT of the urease genes"
 FT primer_bind
 FT complement (46..67)
 FT /tag= c
 FT /note= "BL1 primer"
 FT 102..818
 FT /tag= d
 FT /product= urease A subunit
 FT CDS
 FT 822..2531
 FT /tag= e
 FT /product= urease B subunit
 FT 2546..2569
 FT primer_bind
 FT 2546..2569
 FT /tag= f
 FT /note= "BL2 primer"
 FT 2693..2735
 FT /tag= g
 FT /note= "T7 terminator"
 FT terminator
 FT /note= "T7 terminator"
 XX
 XX WO963732-A1.
 XX
 XX 31-OCT-1996.
 XX
 XX 25-APR-1996; 96WO-US05800.
 XX
 XX 06-DEC-1995; 95US-0568122.
 XX 28-APR-1995; 95US-0431041.
 XX
 XX (ORAV-) ORAVAX INC.
 XX
 XX Ackerman SK, Bhagat H, Ermak T, Guirakhoo F, Kleantous H;
 PI Lee CK, Monath TP, Pappo J, Soman G, Sussman I;
 PI Thomas WD, Weltzin RA;
 XX
 XX WPI: 1996-497373/49.
 DR P-PSDB; AAM07193;
 DR P-PSDB; AAM07194.
 XX
 XX Vaccine for inducing mucosal response to Helicobacter - contg.
 PT multimeric urease complex and pref. an antibiotic, anti-secretory
 PT agent or bismuth salt
 XX
 XX
 PS Disclosure: Page 68-70; 98pp; English.
 XX
 XX A cDNA clone (AAT44351), derived from PORV214, includes the ureA and
 CC ureB sequences coding for the urease A (AAM07193) and urease B
 CC (AAM07194) subunits of Helicobacter pylori clinical isolate CPM630.
 CC To obtain PORV214, a genomic DNA library of CPM630 was screened
 CC with anti-Helicobacter urease antibody. A 17 kb SalI fragment from
 CC an isolated clone was subcloned into pUC18 to give pSCPI. PCR
 CC primers (AAT44352-53) were used to amplify a 2.5 kb fragment from
 CC pSCPI, which was inserted into PET24+ to give PORV214. The vector
 CC was utilised in the produ. of recombinant, enzymatically inactive,
 CC multimeric urease in E. coli transformants for use in vaccines to
 CC treat or prevent Helicobacter infection.

SQ Sequence 2735 BP; 875 A; 562 C; 635 G; 663 T; 0 other;
 Query Match 7.6%; Score 61; DB 17; Length 2735;
 Best Local Similarity 100.0%; Pred. No. 2.8e-20;
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 744 gaattcgagctcgctgcacaaagcttgcgcgcactgcagaccacaccacacactg 803
 Db 2564 gaattcgagctcgctgcacaaagcttgcgcgcactgcagaccacaccacacactg 2623
 QY 804 a 804
 Db 2624 a 2624
 RESULT 9
 AAT08489
 ID AAT08489 standard; DNA; 5248 BP.
 XX
 AC AAT08489;
 XX
 DT 15-MAR-1996 (first entry)
 XX
 DE Plasmid PET Trc S05/N1 containing SGI.1 scFv.
 XX
 KW Complement C5; haemolysis; kidney; glomerulonephritis;
 KW monoclonal antibody; antiinflammatory; antibody engineering;
 KW humanised antibody; complementarity determining region; CDR;
 KW single chain antibody; scFv; vector; PET Trc S05/N1;
 KW Escherichia coli; ds.
 XX
 OS Synthetic.
 XX
 XX WO9529697-A1.
 XX
 XX 09-NOV-1995.
 XX
 XX 01-MAY-1995; 95WO-US05688.
 XX
 XX 02-MAY-1994; 94US-0236208.
 XX
 XX (ALEX-) ALEXION PHARM INC.
 XX
 XX Evans MJ, Malis L, Mueller EE, Nye SH, Rollins S;
 PI Rother RP, Springhorn J P, Squinto SP, Thomas TC;
 PI Wang T, Wilkins JA;
 XX
 XX WPI: 1995-392923/50.
 DR
 XX
 PT Treating glomerulonephritis with antibody against complement C5
 PT component - to inhibit complement induced cell lysis
 XX
 XX Example 11; Page 141-146; 181pp; English.
 XX
 XX A DNA construct (AAT08488) coding for a humanised CDR-grafted
 CC scFv, designated SGI.1 scFv D012 (AAR77616), was subcloned into
 CC bacterial expression vector PET Trc S05/N1 and the resulting
 CC plasmid (sequence given in AAT08489) was transformed into
 CC Escherichia coli W61. Humanised, recombinant scFv was obtd.
 CC which retained the ability of Mab SGI.1 to block human complement
 CC C3a and C5b-9 generation and which was thus able to reduce
 CC glomerular inflammation and kidney dysfunction associated with
 CC glomerulonephritis.
 XX
 SQ Sequence 5248 BP; 1226 A; 1373 C; 1381 G; 1268 T; 0 other;
 Query Match 7.6%; Score 61; DB 16; Length 5248;
 Best Local Similarity 100.0%; Pred. No. 2.6e-20;
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 744 gaattcgagctcgctgcacaaagcttgcgcgcactgcagaccacaccacacactg 803

DB 5052 gaattcgagctcgctgcagcaagcttgcgcgcactcgagcaccaccaccaccactg 5111
 QY 804 a 804
 DB 5112 a 5112

RESULT 10
 AAT41898
 ID AAT41898 standard; DNA: 5248 BP.

AC AAT41898;
 XX
 DT 01-FEB-1997 (first entry)
 XX
 DE Vector pET Trc SOS/N1.
 XX
 KM Vector; PET Trc SOS/N1; myelin basic protein; MBP;
 KM proteolipid protein; PLP; multiple sclerosis; autoimmune disease;
 KM diagnosis; therapy; T-lymphocyte; T-cell; anergy; apoptosis; ds;
 KM cyclic.

XX Synthetic.
 OS
 PN WO9634622-A1.

PD 07-NOV-1996.
 XX
 PF 22-APR-1996; 96WO-US05611.

XX 07-JUN-1995; 95US-0482114.
 PR 02-MAY-1995; 95US-0431644.
 PR 02-MAY-1995; 95US-0431648.

XX (ALEX-) ALEXION PHARM INC.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Leonardo MJ, Matlis L, McFarland HP, Mueller EE, Mueller JP;
 PI Nye SH, Pelfrey CM, Squinto SP, Wilkins JA;

PI WPI; 1996-505898/50.

PT New human myelin basic protein and proteolipid protein variant(s)
 PT used in the assessment, diagnosis and treatment of multiple
 PT sclerosis

PS Disclosure; Page 100-103; 156pp; English.

CC Plasmid PET Trc SOS/N1 (AAT41898) is a bacterial expression vector.
 CC It is suitable for expression of human native and modified myelin
 CC basic protein and proteolipid protein (see also AAM00399-400 and
 CC AAM06101-08) in bacterial host cells, e.g. Escherichia coli.

SQ Sequence 5248 BP; 1226 A; 1373 C; 1381 G; 1268 T; 0 other;

Query Match 7.6%; Score 61; DB 17; Length 5248;
 Best Local Similarity 100.0%; Pred. No. 2.6e-20;
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 744 gaattcgagctcgctgcagcaagcttgcgcgcactcgagcaccaccaccaccactg 803
 DB 5052 gaattcgagctcgctgcagcaagcttgcgcgcactcgagcaccaccaccaccactg 5111

QY 804 a 804
 DB 5112 a 5112

RESULT 11
 AAX28091
 ID AAX28091 standard; DNA: 6363 BP.

XX AAX28091;
 AC
 XX
 DT 10-JUN-1999 (first entry)
 XX
 DE Plasmid pACYC-pol.
 XX
 KM MULV reverse transcriptase; dsDNA production; cDNA library production;
 KM ss.

OS Synthetic.
 XX
 PN US5891637-A.

PD 06-APR-1999.
 XX
 PF 15-SEP-1997; 97US-0929967.

XX 15-OCT-1996; 96US-0732861.
 PR 15-SEP-1997; 97US-0929967.

XX (GETH) GENENTECH INC.

PI Ruppert SJ;

PI WPI; 1999-253852/21.

PT Producing double stranded cDNA molecules from mRNA transcripts and a
 PT novel xenotropic murine leukemia virus reverse transcriptase -
 PT useful for constructing cDNA libraries

PS Example 5; Fig 11; 70pp; English.

XX This sequence represents a plasmid used to test the method of
 CC the invention. The method is for the production of double
 CC stranded cDNA (dsDNA) molecules from mRNA transcripts and a xenotropic
 CC murine leukaemia virus reverse transcriptase (XV-MuLV). The method, the
 CC host cells, reverse transcriptases and nucleic acids disclosed may be
 CC used to produce high quality, normalized, full-length, directionally
 CC cloned, cDNA libraries. Use of the method to construct cDNA libraries
 CC ensures that the reverse transcriptase is able extend the first strand of
 CC the cDNA up to the ultimate 5' end of the mRNA, therefore ensuring that
 CC all sequences are represented in the library, whereas previously, some
 CC sequences would be lost due to incomplete transcription (e.g. as a result
 CC of hairpin loop formation) or mRNA degradation (e.g. as a result of RNase
 CC H activity). Also, carrying the process out in cells means that enzymes
 CC and proteins present in the cell will repair any mistakes or nicks in the
 CC product DNA. Modifying the mRNA for insertion into a vector by adding the
 CC 5' cap provides a convenient way of priming it for cDNA synthesis. Also,
 CC the method provides a fast and clean way of synthesizing cDNA
 CC clones.

SQ Sequence 6363 BP; 1576 A; 1701 C; 1645 G; 1441 T; 0 other;

Query Match 6.8%; Score 55; DB 20; Length 6363;
 Best Local Similarity 100.0%; Pred. No. 2.2e-17;
 Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 gagcttcgctgcagcaagcttgcgcgcactcgagcaccaccaccaccactga 804
 DB 4656 gagcttcgctgcagcaagcttgcgcgcactcgagcaccaccaccaccaccactga 4710

RESULT 12
 AAF74635
 ID AAF74635 standard; DNA: 59 BP.
 XX
 AC AAF74635;
 XX
 DT 15-MAY-2001 (first entry)
 XX

XX Cloning vector pTK7 nucleotide fragment SEQ ID NO:9.

XX	Nicking endonuclease; N.BstNBI; restriction endonuclease;
KW	non-thio strand displacement amplification; PfuI modification methylase;
KM	ds.
XX	
OS	Bacillus stearoerophilus.
XX	
PN	US6191267-B1.
XX	
PD	20-FEB-2001.
XX	
PF	02-JUN-2000; 2000US-0586935.
XX	
PR	02-JUN-2000; 2000US-0586935.
XX	
PA	(NEW) NEW ENGLAND BIOLABS INC.
XX	
PI	Kong H, Higgins LS, Dalton M, Kucera RB, Schildkraut I;
XX	
DR	WFI: 2001-210381/21.
XX	
PT	Recombinant DNA encoding a novel nicking endonuclease, N.BstNBI useful
PX	in non-thio strand displacement amplification
PS	
XX	Disclosure; Fig 6; 30pp; English.
CC	The present invention describes an isolated DNA (I) coding for the
CC	N.BstNBI restriction endonuclease obtainable from ATCC Accession
CC	No. PTA-1945. Also described are: (1) a vector (II) comprising (I);
CC	(2) a host cell transformed by (II); and (3) producing an N.BstNBI
CC	restriction endonuclease by culturing a host cell transformed with
CC	(II) under conditions suitable for expression of the endonuclease.
CC	(I) is useful for producing N.BstNBI nicking endonuclease, which
CC	is useful in non-thio strand displacement amplification. The present
CC	sequence represents a fragment from the pKMT7 cloning vector,
CC	which is given in the exemplification of the present invention.
SQ	
Sequence	59 BP; 11 A; 22 C; 17 G; 9 T; 0 other:
Query Match	6.2%; Score 50; DB 22; Length 59;
Best Local Similarity	100.0%; Fred. No. 8.3e-15;
Matches	50; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
QY	741 tcgcgaattcgagctccgctgcgacaagtcttgcgccgacactcgcagaccacc 790
Db	10 tccgattcgagctccgctgcgacaagtcttgcgccgacactcgcagaccacc 59
RESULT	13
AAA61167	
ID	AAA61167 standard; DNA; 357 BP.
XX	
AC	AAA61167;
XX	
DT	12-OCT-2000 (first entry)
XX	
DE	SEN virus genome fragment SEQ ID NO: 137.
XX	
KM	SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease;
KW	proliferative disorder; hepatopathy; hepatitis; viral infection;
KM	vaccination; gene therapy; ds.
XX	
OS	Hepatitis virus.
XX	
FN	WO200028039-A2.
XX	
PD	18-MAY-2000.
XX	
PF	09-NOV-1999; 99WO-EP08566.
XX	
R	10-NOV-1998; 98IT-MIO2437.
RR	30-APR-1999; 99IT-MIO0923.

PR	14-MAY-1999;	99EP-0830298.
PR	16-JUL-1999;	99EP-0113932.
XX	(DIAS-) DIASORIN SRL.	
PA	Primi D, Fiordalisi G,	Mantero GL, Mattioli S, Sottini A;
PI	Bonelli F, Vaglini L,	Oliviero P, Dal Corso A, Bonelli M;
PI		
XX	WPI; 2000-376551/32.	
XX		
XX	Nucleic acids representing the genome of the SEN virus (SENV) and	
PT	encoded proteins, useful for treatment of hepatopathies, inflammatory	
PT	diseases and proliferative disorders such as cancer -	
XX		
PS	Example 25; Page 97; 392pp; English.	
CC	The present invention is concerned with the sequence of the genome of the	
CC	SEN virus (SENV), and the proteins encoded by it. SENV is thought to be	
CC	the cause of hepatopathies, which are not linked to the presence of the	
CC	hepatitis A, B and E viruses in man. The genome and proteins of this	
CC	virus can be used in gene therapy and vaccination against the virus,	
CC	which also causes disorders of the gastrointestinal tract, including	
CC	Crohn's disease and lupus erythematosus, inflammatory diseases, and	
CC	proliferative disorders such as cancer.	
XX		
SQ	Sequence 357 BP; 117 A; 109 C; 76 G; 55 T; 0 other:	
	Query Match	5.2%; Score 42; DB 21; Length 357;
	Best Local Similarity	100.0%; Pred. No. 5.9e-11;
	Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
OY	763 aaagctgcgcccacatcgagcaccacccacacactga 804	
Dd	316 aagctgcgcccacatcgagcaccacccacacactga 357	
	RESULT 14	
	AAA61166	
ID	AAA61166 standard; DNA; 586 BP.	
XX		
XX	AAA61166;	
DT	12-OCT-2000 (first entry)	
DE		
XX	SEN virus genome fragment SEQ ID NO: 133.	
KW	SEN virus; SENV: gastrointestinal tract disorder; inflammatory disease;	
KW	proliferative disorder; hepatopathy; hepatitis; viral infection;	
KW	vaccination; gene therapy; ds.	
OS	Hepatitis virus.	
PN	WO200028039-A2.	
PD	18-MAY-2000.	
PE	09-NOV-1999; 99WO-EP08566.	
PR	10-NOV-1998; 98IT-MI02437.	
PR	30-APR-1999; 99IT-MI00923.	
PR	14-MAY-1999; 99EP-0830298.	
PR	16-JUL-1999; 99EP-0113932.	
PA	(DIAS-) DIASORIN SRL.	
PI	Primi D, Fiordalisi G,	Mantero GL, Mattioli S, Sottini A;
PI	Bonelli F, Vaglini L,	Oliviero P, Dal Corso A, Bonelli M;
DR	WPI; 2000-376551/32.	
XX		
XX	Nucleic acids representing the genome of the SEN virus (SENV) and	
XX	encoded proteins, useful for treatment of hepatopathies, inflammatory	

PT diseases and proliferative disorders such as cancer -
XX
PS Example 24; Page 94; 392pp; English.

CC The present invention is concerned with the sequence of the genome of the
CC SEN virus (SENV), and the proteins encoded by it. SENV is thought to be
CC the cause of hepatopathies which are not linked to the presence of the
CC hepatitis A, B and E viruses in man. The genome and proteins of this
CC virus can be used in gene therapy and vaccination against the virus,
CC which also causes disorders of the gastrointestinal tract, including
CC Crohn's disease and lupus erythematosus, inflammatory diseases, and
CC proliferative disorders such as cancer.

SO Sequence 586 BP; 155 A; 172 C; 139 G; 119 T; 1 other;

Query Match 5.2%; Score 42; DB 21; Length 586;
Best Local Similarity 100.0%; Pred. No. 5.7e-11;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aagcttgcggcgccactcgagcaccaccaccaccactga 804
|||||
DB 545 aagcttgcggcgccactcgagcaccaccaccaccactga 586

RESULT 15

ID AAX29726 standard; DNA; 1186 BP.

AC AAX29726;

DT 08-JUN-1999 (first entry)

DE Insert from pET21C-clone 2 from MSRV-1.

KW Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
KM rheumatoid polyarthritis; ss.

OS Multiple sclerosis related virus type 1.

PN FR2765588-A1.

PD 08-JAN-1999.

PF 07-JUL-1997; 97FR-0008816.

PR 07-JUL-1997; 97FR-0008816.

PA (INMR) BIO MERIEUX.

DR WPI; 1999-098275/09.

DR P-PSDB; AAW99558.

PT Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT multiple sclerosis or rheumatoid polyarthritis

PS Disclosure: Fig 8; 83pp; French.

CC This sequence represents the insert found in the plasmid pET21C-clone 2
CC containing a fragment from a novel multiple sclerosis related virus
CC type 1 (MSRV1). The sequence can be used in diagnostic, prophylactic or
CC therapeutic compositions to inhibit expression of a multiple sclerosis
CC related virus and/or virus associated with rheumatoid polyarthritis.

SO Sequence 1186 BP; 378 A; 284 C; 260 G; 264 T; 0 other;

DB 11096 aagcttgcggcgccactcgagcaccaccaccaccaccactga 1137

Search completed: August 23, 2001, 15:01:33
Job time: 9851 sec

Query Match 5.2%; Score 42; DB 20; Length 1186;

Best Local Similarity 100.0%; Pred. No. 5.5e-11;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aagcttgcggcgccactcgagcaccaccaccaccactga 804
|||||

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 14:09:06 ; Search time 1867.27 Seconds

(without alignments)
4070.164 Million cell updates/sec

Title: us-09-654-652a-5

Perfect score: 804

Sequence: 1 atggttagcgcaaggattt.....accacacacacacacactga 804

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size: 0

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	22	2.7	154	153	BG412308	BG412308 OV2_40.D0
2	22	2.7	163	173	BG054383	BG054383 OV2_3-G03
3	22	2.7	290	148	BF20855	BF20855 OV1_12-A0
4	22	2.7	380	114	AM287483	AM287483 LG1_228-E
5	22	2.7	476	152	BG322490	BG322490 EML_28-F0
6	22	2.7	509	104	AT948373	AT948373 603044A01
7	22	2.7	510	156	D15218	D15218 R1CC0286A R
8	22	2.7	544	137	BE594445	BE594445 P11_33-E0
9	22	2.7	548	154	BG487461	BG487461 EML_65-C1
10	22	2.7	555	154	BG487511	BG487511 EML_65-H1
11	22	2.7	581	137	BE598959	BE598959 P11_84-D0
12	22	2.7	584	137	BE600643	BE600643 P11_89-F0
13	22	2.7	615	137	BE598799	BE598799 P11_82-A1
14	22	2.7	618	166	BE357292	BE357292 DGI_148-C
15	21	2.6	272	165	BG550683	BG550683 sad22f12.
16	21	2.6	272	160	BBS37820	BBS37820 BBS37820
17	21	2.6	292	162	BE000537	BE000537 RC3-BN007
18	21	2.6	300	151	BE597387	BE597387 su97b10.Y
19	21	2.6	309	160	BBS40994	BBS40994 BBS40994
20	21	2.6	311	128	BB218522	BB218522 BB218522
21	21	2.6	337	31	AV630939	AV630939 AV630939
22	21	2.6	385	121	AM838842	AM838842 CM0-LT005
23	21	2.6	394	158	H49830	H49830 yq21908.r1
24	21	2.6	406	30	AV426195	AV426195 AV426195
25	21	2.6	413	22	A1619419	A1619419 CEST2000
26	21	2.6	430	244	A2472019	A2472019 IM0286020
27	21	2.6	440	107	AU015444	AU015444 AU015444
28	21	2.6	441	219	CNS01P0U	AL155086 Anopheles
29	21	2.6	473	244	AZ470822	AZ470822 IM0285C13
30	21	2.6	480	245	AZ498394	AZ498394 IM0335J10
31	21	2.6	483	250	AZ853499	AZ853499 ZM0156A16
32	21	2.6	487	107	AU084357	AU084357 AU084357
33	21	2.6	494	231	AO655902	AO655902 Sheared D
34	21	2.6	499	243	AZ392169	AZ392169 IM0154K15
35	21	2.6	507	164	BE212984	BE212984 IPBn0133
36	21	2.6	534	241	AZ287683	AZ287683 RPT-23-1
37	21	2.6	546	228	AQ439343	AQ439343 HS-1060-B
38	21	2.6	556	247	AZ642854	AZ642854 IM0506M06
39	21	2.6	560	240	AZ261227	AZ261227 RPT-23-4
40	21	2.6	614	111	AM077378	AM077378 f134d10.Y
41	21	2.6	633	238	AZ114598	AZ114598 RPT-23-4
42	21	2.6	641	136	BE469249	BE469249 IPHAK0151
43	21	2.6	651	242	AZ366869	AZ366869 IM0116J15
44	21	2.6	678	241	AZ328225	AZ328225 IM0051F24
45	21	2.6	685	156	C94209	C94209 C94209 Dict

ALIGNMENTS

RESULT	LOCUS	DEFINITION	SEQUENCE	EST	LOCUS	DEFINITION	SEQUENCE
BG412308	154 bp	mRNA	OV2_40.D02.bl_A002 Ovary 2 (OV2)	13-MAR-2001	BG412308	13-MAR-2001	EST
BG412308	154 bp	mRNA	OV2_40.D02.bl_A002 Ovary 2 (OV2)	13-MAR-2001	BG412308	13-MAR-2001	EST
VERSION	BG412308.1	GI:13317861			VERSION	BG412308.1	GI:13317861
KEYWORDS	EST				KEYWORDS	EST	
SOURCE	Sorghum				SOURCE	Sorghum	
ORGANISM	Sorghum bicolor				ORGANISM	Sorghum bicolor	
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC				REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC	
AUTHORS	1 (bases 1 to 154)				AUTHORS	1 (bases 1 to 154)	
TITLE	Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt				TITLE	Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt	
	An EST database from Sorghum: ovaries of varying immature stages					An EST database from Sorghum: ovaries of varying immature stages	

JOURNAL COMMENT

Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Seq primer: JEN REV
 High quality sequence stop: 142
 POLYA-No.

FEATURES

Location/Qualifiers
 1..154
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Ovary 2 (OV2)"
 /note="Organ: Mix of ovaries of varying immature stages from 8-week-old plants; Vector: pBluescript II from lambda Zap II; Site: 1: XhoI; Site: 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
 BASE COUNT 30 a 50 c 60 g 14 t

ORIGIN

Query Match 2.7%; Score 22; DB 153; Length 154;
 Best Local Similarity 100.0%; Pred. No. 0.43;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy

780 cgagcaccacccacccacac 801
 ||||||||||||||||||||
 65 CGAGCACCCACCCACCCAC 86

RESULT 2

BG054383 163 bp mRNA EST 25-JAN-2001
 LOCUS OV2_3.G03.bl_A002 Ovary 2 (OV2) Sorghum bicolor cDNA, mRNA
 DEFINITION sequence.
 ACCESSION BG054383
 VERSION BG054383.1 GI:12511071
 KEYWORDS EST.
 SOURCE Sorghum bicolor
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade: Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 163)
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 L.H.
 An EST database from Sorghum: ovaries of varying immature stages
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Seq primer: JEN REV
 High quality sequence stop: 160
 POLYA-No.

REFERENCE

1 (bases 1 to 163)
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 L.H.
 An EST database from Sorghum: ovaries of varying immature stages
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Seq primer: JEN REV
 High quality sequence stop: 160
 POLYA-No.

TITLE

Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Seq primer: JEN REV
 High quality sequence stop: 160
 POLYA-No.

JOURNAL COMMENT

Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Seq primer: JEN REV
 High quality sequence stop: 160
 POLYA-No.

FEATURES

Location/Qualifiers
 1..163
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Ovary 2 (OV2)"
 /note="Organ: Mix of ovaries of varying immature stages from 8-week-old plants; Vector: pBluescript II from lambda Zap II; Site: 1: XhoI; Site: 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
 BASE COUNT 30 a 61 c 54 g 18 t

ORIGIN

Query Match 2.7%; Score 22; DB 173; Length 163;
Best Local Similarity 100.0%; Pred. No. 0.43;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 780 cgaagcaccaccaccaccaccac 801
|||||
Db 93 CGAGCACACCACCACCACCAC 114

RESULT 3

BF420855 290 bp mRNA EST 28-NOV-2000
LOCUS OVL_12_A02_b1_A002 Ovary 1 (OVL) Sorghum bicolor cDNA, mRNA
DEFINITION sequence.

ACCESSION BF420855
VERSION BF420855.1 GI:11408844

KEYWORDS

SOURCE sorghum.
ORGANISM Sorghum bicolor

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 290)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.

AUTHORS

REFERENCE An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 140
POLYA-No.

FEATURES

source Location/Qualifiers

1..290
/organism="Sorghum bicolor"
/db_xref="taxon:4558"

/clone_11b="Ovary 1 (OVL)"
/note="Organ: Mix of ovaries of varying immature stages
from 8-week-old plants; Vector: pBluescript II from lambda
zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda zap II.
Clones to be sequenced were prepared by mass excision."

BASE COUNT 54 a 107 c 85 g 42 t 2 others

ORIGIN

Query Match 2.7%; Score 22; DB 148; Length 290;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 780 cgaagcaccaccaccaccaccac 801
|||||
Db 104 CGAGCACACCACCACCACCAC 125

RESULT 4

AM287483 380 bp mRNA EST 19-JUL-2000
LOCUS LGL_228_E07_b1_A002 Light Grown 1 (LGL) Sorghum bicolor cDNA, mRNA
DEFINITION sequence.
ACCESSION AM287483
VERSION AM287483.2 GI:6859477

KEYWORDS

EST.
Sorghum.
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 380)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.

REFERENCE An EST database from Sorghum: light-grown seedlings
JOURNAL Unpublished (2000)
COMMENT On Jan 6, 2000 this sequence version replaced gi:6677327.
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 379
POLYA-No.

FEATURES

source Location/Qualifiers

1..380
/organism="Sorghum bicolor"
/db_xref="taxon:4558"

/clone_11b="Light Grown 1 (LGL)"
/note="Organ: 10- to 14-day-old light-grown (greenhouse)
seedlings; Vector: Lambda zap; Site_1: XhoI; Site_2: EcoRI
; The library was made from poly-A RNA in the cloning
vector lambda zap II. Clones to be sequenced were
prepared by mass excision."

BASE COUNT 79 a 145 c 104 g 52 t

ORIGIN

Query Match 2.7%; Score 22; DB 114; Length 380;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 780 cgaagcaccaccaccaccaccac 801
|||||
Db 110 CGAGCACACCACCACCACCAC 131

RESULT 5

BG322490 476 bp mRNA EST 27-FEB-2001
LOCUS EML_28_F03_b1_A002 Embryo 1 (EML) Sorghum bicolor cDNA, mRNA
DEFINITION sequence.

ACCESSION BG322490
VERSION BG322490.1 GI:13152168

KEYWORDS

SOURCE sorghum.
ORGANISM Sorghum bicolor

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 476)
Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.

AUTORS An EST database from Sorghum: developing embryos
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence

is 20.
Seq primer: JEN REV
High quality sequence stop: 393
POLYA-No.

FEATURES
source location/Qualifiers

1..476
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 99 a 177 c 127 g 73 t

Query Match 2.7%; Score 22; DB 152; Length 476;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 cgagcaccaccaccaccac 801
|||||
Db 104 CGAGCACCCACCCACCCAC 125

RESULT 6
LOCUS A1948373 509 bp mRNA EST 19-AUG-1999
DEFINITION 603044A01.xl 603 - stressed root cDNA library from Wang/Bohnert lab
ACCESSION A1948373
VERSION A1948373.1 GI:5740683
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 509)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 603044 row: A column: 01.
FEATURES
source location/Qualifiers

1..509
/organism="Zea mays"
/cultivar="873"
/db_xref="taxon:4577"
/clone_lib="603 - stressed root cDNA library from Wang/Bohnert lab"
/tissue_type="seedling"
/dev_stage="salt stress"
/lab_host="E. coli XL Gold"
/note="Organ: root; Vector: pBluescriptII SK(+) XR; Seedling stressed root cDNA library from Wang/Bohnert lab"

BASE COUNT 122 a 111 c 131 g 145 t
ORIGIN

Query Match 2.7%; Score 22; DB 104; Length 509;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 cgagcaccaccaccaccac 801
|||||
Db 248 CGAGCACCCACCCACCCAC 227

RESULT 7
LOCUS D15218/c 510 bp mRNA EST 08-JUL-1999
DEFINITION RICC0286A Rice callus Oryza sativa subsp. japonica cDNA clone C0286A, mRNA sequence.
ACCESSION D15218
VERSION D15218.1 GI:286410
KEYWORDS EST.
SOURCE Oryza sativa subsp. japonica.
ORGANISM Oryza sativa subsp. japonica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Erihartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 510)
AUTHORS Sasaki,T. and Minobe,Y.
TITLE Rice cDNA from callus
JOURNAL Unpublished (1994)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@agr.affrc.go.jp
PROJECT "RGP".
Seq primer: oligo(dT).
FEATURES
source location/Qualifiers

1..510
/organism="Oryza sativa subsp. japonica"
/strain="cultivar Nipponbare, sub-species japonica"
/db_xref="taxon:39947"
/clone="C0286A"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2: NotI; cDNA prepared from rice callus mRNAs by using oligo(dT) as a primer and ligating to the SalI-NotI site of pBluescript II SK+ phagemid."
BASE COUNT 80 a 121 c 171 g 106 t 32 others
ORIGIN

Query Match 2.7%; Score 22; DB 156; Length 510;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 cgagcaccaccaccaccac 801
|||||
Db 174 CGAGCACCCACCCACCCAC 153

RESULT 8
LOCUS BE594445 544 bp mRNA EST 18-AUG-2000
DEFINITION P11_31_E02.D1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BE594445
VERSION BE594445.1 GI:9849518
KEYWORDS EST.
SOURCE Sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 544)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt,L.H.

TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 481
POLYA-No.

FEATURES
source Location/Qualifiers

1..544
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Pathogen induced 1 (PI1)"
/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pBluescript II from lambda Zap II; Site.1: XhoI; Site.2: EcoRI; Two-week-old sorghum plants (BTX 623 cultivar) were infected with pathogen (isolate FRM41 of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptibility to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."

BASE COUNT 123 a 197 c 141 g 83 t
ORIGIN

Query Match 2.7%; Score 22; DB 137; Length 544;
Best Local Similarity 100.0%; Pred. No. 0.47; Mismatches 0; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 780 cgaagcaccaccaccaccacac 801
|||||
DB 93 CGAGCACACACACACACAC 114

RESULT 9
BG487461 548 bp mRNA EST 27-MAR-2001
LOCUS EML_65.C11.b1_A002 Embryo 1 (EM1) Sorghum bicolor CDNA, mRNA
DEFINITION
ACCESSION BG487461
VERSION BG487461.1 GI:13468696
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade: Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 548)
Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
An EST database from Sorghum: developing embryos
Unpublished (2000)
JOURNAL
TITLE
AUTHORS
COMMENT

The University of Georgia
Department of Botany
Contact: Cordonnier-Pratt MM
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805

Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 546
POLYA-No.

FEATURES
source Location/Qualifiers

1..548
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr; Vector: pBluescript II from lambda Zap II; Site.1: XhoI; Site.2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 121 a 194 c 137 g 96 t
ORIGIN

Query Match 2.7%; Score 22; DB 154; Length 548;
Best Local Similarity 100.0%; Pred. No. 0.47; Mismatches 0; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 780 cgaagcaccaccaccaccacac 801
|||||
DB 31 CGAGCACACACACACACAC 52

RESULT 10
BG487511 555 bp mRNA EST 27-MAR-2001
LOCUS EML_65.H11.b1_A002 Embryo 1 (EM1) Sorghum bicolor CDNA, mRNA
DEFINITION
ACCESSION BG487511
VERSION BG487511.1 GI:13468746
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade: Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 555)
Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
An EST database from Sorghum: developing embryos
Unpublished (2000)
JOURNAL
TITLE
AUTHORS
COMMENT

The University of Georgia
Department of Botany
Contact: Cordonnier-Pratt MM
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805

Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 476
POLYA-No.

FEATURES
source Location/Qualifiers

1..555
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr; Vector: pBluescript II from lambda Zap II; Site.1: XhoI; Site.2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 123 a 203 c 142 g 87 t
ORIGIN

Query Match	2.7%	Score 22;	DB 154;	Length 555;
Best Local Similarity	100.0%	Pred. No. 0.47;		
Matches 22:	Conservative	0;	Mismatches	0;
			Indels;	0;
			Gaps	0;
Oy	780	CGAGGACCAACCAACCAACCAAC	801	
Db	104	CGAGGACCAACCAACCAACCAAC	125	

RESULT	11
BE598959	
LOCUS	581 bp mRNA EST 18-Aug-2000
DEFINITION	P11.84.D06.b1.A02 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
ACCESSION	mRNA sequence.
VERSION	BE598959.1 GI:9854018
KEYWORDS	EST.
SOURCE	sorghum.

ORGANISM	TITLE	JOURNAL	COMMENT
Sorghum bicolor Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum. 1 (bases 1 to 581) Cordonnier-Pratt, M.-M., Gingle, A., Dean, R., Sudman, M. and Pratt , L.H.	An EST database from sorghum: pathogen-induced plants unpublished (2000)		MM

Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@atluga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 542
PolyA-No.

FEATURES	SOURCE
Location/Qualifiers	1. 581
/organism="Sorghum bicolor"	
/db_xref="taxon:4558"	
/clone_lib="Pathogen induced 1 (PI1)"	
/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pBluescript II from lambda Zap II; Site:1: XhoI; Site:2: EcoRI; Two-week-old sorghum plants (BRX 623 cultivar) were infected with pathogen (isolate FRM421 of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings 48 hours after inoculation. Note: Young seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptibility to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."	
BASE COUNT	125 a 209 c 145 g 102 t
ORIGIN	

```

Query Match      2.7%; Score 22; DB 137; Length 581;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 780 cgagcaccacacacacacac 801
|||||

```

Db 23 CGAGCACCACCACCACCAC 44

RESULT	12
BE600643	
LOCUS	
DEFINITION	BE600643 584 bp mRNA EST 18-AUG-2000 P.tl1.89.F08.b1.A002 Pathogen induced 1 (PtI) Sorghum bicolor CDNA.
ACCESSION	mRNA sequence.
VERSION	BE600643
KEYWORDS	BE600643.1 GI:9855716
SOURCE	EST. Sorghum.

ORGANISM	TITLE	JOURNAL	COMMENT
Sorghum bicolor	An EST database from sorghum: pathogen-induced plants	Unpublished (2000)	Contact: Cordonnier-Pratt MM
Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta			
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC			
clade; Panicoideae; Andropogoneae; Sorghum.			
1 (bases 1 to 584)			
REFERENCE			
AUTHORS			
Cordonnier-Pratt, M.-M., Gingle, A., Dean, R., Sudman, M. and Pratt, L. H.			

Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel.: 706 542 1860
Fax: 706 542 1805
Email: mpirattetuga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 568
POLYA=No.

```

FEATURES
source
Location/Qualifiers
1..584
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_id="Pathogen induced 1 (PII)"
/note="Organ: Anthracnose-infected leaves from
two-week-old sorghum plants 48 hr after inoculation;
Vector: pBluescript II from lambda Zap II; Site1: XhoI
Site2: EcoRI; Two-week-old sorghum plants (BRX 623
cultivar) were infected with pathogen (isolate FRM421 o
Colletotrichum graminicola, which is a sorghum isolate)
RNA was prepared from infected leaves harvested from 45
seedlings (2 weeks old) exhibit juvenile resistant
reaction, which is an incompatible interaction. As they
grow older (4 weeks or older), plants resume susceptibl
to anthracnose disease. The library was made from poly-
RNA in the cloning vector lambda Zap II. Clones to be
sequenced were prepared by mass excision. WARNING: While
most or all ESTs are expected to derive from the host
plant, no effort was made to eliminate ESTs deriving fr
the pathogen."
130 a 211 c 149 g 94 t

```

Query Match	2.7%:	Score 22;	DB 137;	Length 584;
Best Local Similarity	100.0%;	Pred. No. 0.47;		
Matches 22; Conservative	0;	Mismatches	0;	Indels 0;
Gaps				
QY 780	cgagcaccaccaccaccaccac	801		
Db 93	CGAGCACCACCACCACCACCAC	114		
RESULT 13				
LOCUS BE598799				
DEFINITION	BE598799	615 bp	mrna	
	P11_g2_A10.01	A002	Pathogen induced 1 (P11)	Sorghum bicolor cDNA
			EST	18-AUG-2000

mRNA sequence.
 BE598799
 BE598799.1 GI:9853872
 EST.
 SOURCE
 ORGANISM
 Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Sorghum.
 1 (bases 1 to 615)
 Cordonier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
 L.H.
 An EST database from Sorghum: pathogen-induced plants
 Unpublished (2000)
 Contact: Cordonier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: JEN REV
 High quality sequence stop: 587
 POLYA=NO.

FEATURES
 source
 Location/Qualifiers
 1..615
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Pathogen induced 1 (PI1)"
 /note="Organ: Anthracnose-infected leaves from
 two-week-old sorghum plants 48 hr after inoculation;
 Vector: pBluescript II from lambda zap II; Site.1: XhoI;
 Site.2: EcoRI; Two-week-old sorghum plants (BRX 623
 cultivar) were infected with pathogen (isolate PM421 of
 Colletotrichum graminicola, which is a sorghum isolate).
 RNA was prepared from infected leaves harvested from 45
 seedlings 48 hours after inoculation. Note: young
 seedlings (2 weeks old) exhibit juvenile resistant
 reaction, which is an incompatible interaction. As they
 grow older (4 weeks or older), plants resume susceptibility
 to anthracnose disease. The library was made from poly-A
 RNA in the cloning vector lambda zap II. Clones to be
 sequenced were prepared by mass excision. WARNING: While
 most or all ESTs are expected to derive from the host
 plant, no effort was made to eliminate ESTs deriving from
 the pathogen."

BASE COUNT
 ORIGIN
 134 a 214 c 164 g 103 t

Query Match 2.7%; Score 22; DB 137; Length 615;
 Best Local Similarity 100.0%; Pred. No. 0.47;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 cgagcaccaccaccaccac 801
 ||||||||||||||||||||
 Db 69 CGAGCACACACACCACCAC 90

RESULT 14
 BE357292
 LOCUS
 DEFINITION
 BE357292 618 bp mRNA EST 20-JUL-2000
 DGI_148.C02.bi_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA
 sequence.
 BE357292
 VERSION
 BE357292.1 GI:92298849
 KEYWORDS
 EST
 SOURCE
 Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Sorghum.
 1 (bases 1 to 618)
 Cordonier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 L.H.
 An EST database from Sorghum: dark-grown seedlings
 Unpublished (2000)
 Contact: Cordonier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: JEN REV
 High quality sequence stop: 617
 POLYA=NO.

FEATURES
 source
 Location/Qualifiers
 1..618
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Dark Grown 1 (DGI)"
 /note="Organ: 5-day-old dark-grown seedlings; Vector:
 Lambda Zap; Site.1: XhoI; Site.2: EcoRI; The library was
 made from poly-A RNA in the cloning vector lambda zap II.
 Clones to be sequenced were prepared by mass excision."

BASE COUNT
 ORIGIN
 134 a 222 c 157 g 105 t

Query Match 2.7%; Score 22; DB 166; Length 618;
 Best Local Similarity 100.0%; Pred. No. 0.47;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 cgagcaccaccaccaccac 801
 ||||||||||||||||||||
 Db 99 CGAGCACACACACCACCAC 120

RESULT 15
 BG550683
 LOCUS
 DEFINITION
 BG550683 272 bp mRNA EST 09-APR-2001
 sad22f12.y1.Gm-cl074 Glycine max cDNA clone GENOME SYSTEMS CLONE
 ID: Gm-cl074-1055 5', mRNA sequence.
 BG550683
 VERSION
 BG550683.1 GI:13562463
 KEYWORDS
 EST
 SOURCE
 soybean.
 ORGANISM
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 272)
 Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,J., Corryell,V., Khanna
 A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
 Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
 R., Riltter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
 R., Waterston,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information

call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Putative full length read
 vector to vector length is 273.

FEATURES

source

1.272
 location/Qualifiers
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl074-1055"
 /clone_lib="Gm-cl074"
 /tissue_type="seedlings induced for HR (hypersensitive response)"
 /dev_stage="9-11 day old"
 /lab_host="DH10B"
 /note="Vector: Bluescript II SK+; Site.1: EcoRI; Site.2: XhoI; The cDNA library was constructed from mRNA isolated from 9-11 day old seedlings that were induced for HR (hypersensitive response) by vacuum infiltrating plant tissue with *Pseudomonas syringae* pv. *glycinea* carrying the *avrB* gene (Genetics 141:1597-1604). Plant tissue (expanded unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36, and 53 hrs after inoculation and their mRNA pooled equally for cDNA construction. The library was prepared using the Stratagene Bluescript II SK(+) library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with an XhoI restriction site. EcoRI adaptors were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA insert is protected from XhoI digestion via methylation during first strand synthesis. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the Bluescript vector. The ligated cDNA fragments were transformed into *E. coli* Electromax DH10B host cells. Plant care, inoculations, and library construction were performed by Steve Clough (Lila Vodkin lab, University of Illinois)."

BASE COUNT 79 a 58 c 55 g 80 t
 ORIGIN

Query Match 2.68; Score 21; DB 155; Length 272;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 783 gcaccaccaccaccaccactg 803
 ||||||||||||||||||||
 Db 11 GCACCACCACCACCACCTG 31

Search completed: August 23, 2001, 14:09:23
 Job time: 8655 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:18:28 ; Search time 19.47 Seconds

(without alignments)
614.030 Million cell updates/sec

Title: US-09-654-652A-3

Perfect score: 349
Sequence: 1 MNKKTKAVKSAALAVAAAAA.....AKGAKVNPNGHKRYRVNFEH 349

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size: 0

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	349	100.0	349	1 GUB_FIBSU	P17989 fibroblacter
2	9	2.6	337	1 HAIR_DROME	P14003 drosophila
3	9	2.6	537	1 MMSA_BOVIN	Q07536 bos taurus
4	9	2.6	918	1 KPCM_MOUSE	Q62101 mus musculu
5	9	2.6	1475	1 N153_HUMAN	P49790 homo sapien
6	8	2.3	37	1 ANP3_PSEAM	P02733 pseudopleur
7	8	2.3	82	1 ANPA_PSEAM	P04002 pseudopleur
8	8	2.3	268	1 ZEST_DROME	Q24597 drosophila
9	8	2.3	317	1 CYSD_RHIME	P36892 rhizobium m
10	8	2.3	317	1 YE80_MYCTU	P71761 mycobacteri
11	8	2.3	368	1 CHEB_PSEAE	Q87125 pseudomonas
12	8	2.3	387	1 SOX1_HUMAN	O00570 homo sapien
13	8	2.3	391	1 HME1_HUMAN	O05925 homo sapien
14	8	2.3	391	1 SOX1_MOUSE	P53783 mus musculu
15	8	2.3	401	1 HME1_MOUSE	P09065 mus musculu
16	8	2.3	407	1 F16P_PEA	P46275 pisum sativ
17	8	2.3	424	1 ULB7_HCMVA	P16770 human cytom
18	8	2.3	428	1 FXB2_MOUSE	Q64733 mus musculu
19	8	2.3	429	1 OORA_MYCTU	O10387 mycobacteri
20	8	2.3	440	1 DCO_DROME	O76324 drosophila
21	8	2.3	458	1 A2AD_HUMAN	P35369 homo sapien
22	8	2.3	459	1 IP3K_RAT	P17105 rattus norv
23	8	2.3	461	1 A2AC_HUMAN	P18825 homo sapien
24	8	2.3	461	1 IP3K_HUMAN	P23677 homo sapien
25	8	2.3	470	1 ESCA_DROME	P25932 drosophila
26	8	2.3	483	1 ELAV_DROME	P16914 drosophila
27	8	2.3	519	1 ELAV_DROVI	P23241 drosophila
28	8	2.3	528	1 TY3H_HUMAN	P07101 homo sapien
29	8	2.3	542	1 CH12_RHIOI	P29027 rhizopus ol
30	8	2.3	559	1 PHF1_MOUSE	Q921b8 mus musculu
31	8	2.3	563	1 ARX_MOUSE	O35085 mus musculu
32	8	2.3	574	1 ZEST_DROME	P09956 drosophila
33	8	2.3	611	1 XINA_PSEFL	P14768 pseudomonas

34	8	2.3	618	1 ZEST_DROVI	Q24762 drosophila
35	8	2.3	652	1 FXO1_MOUSE	Q9160 mus musculu
36	8	2.3	655	1 FXO1_HUMAN	Q12778 homo sapien
37	8	2.3	666	1 HNF4_DROME	P49866 drosophila
38	8	2.3	668	1 PAB5_ARATH	Q05196 arabidopsis
39	8	2.3	732	1 POK_DROME	Q01842 drosophila
40	8	2.3	741	1 ERF2_PICPI	P23637 picchia pinu
41	8	2.3	766	1 EYA_DROME	Q05201 drosophila
42	8	2.3	801	1 ZFY_HUMAN	P08048 homo sapien
43	8	2.3	828	1 NAPA_ECOLI	P33937 escherichia
44	8	2.3	880	1 BRCA_DROME	Q24206 drosophila
45	8	2.3	1077	1 HIES_DROME	Q02308 drosophila

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	349 AA.
GUB_FIBSU				
AC	P17989:			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)			
DE	(1,3-1,4-BETA-D-GLUCAN-4-GLUCANOHYDROLASE) (MIXED LINKAGE BETA-GLUCANASE) (LICHENASE).			
OS	Fibroblacter succinogenes (Bacteroides succinogenes).			
CC	Bacteriata; Fibrobacter/Acidobacteria group; Fibrobacter group;			
CC	Fibrobacter.			
OX	NCBI_TaxID=833;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 28-57.			
RC	STRAIN-ISOLATE S85;			
RX	MEDLINE=90299807; PubMed=2193918;			
RA	Teacher R.W., Ertle J.D.;			
RT	"DNA sequence of a fibrobacter succinogenes mixed-linkage beta-glucanase (1,3-1,4-beta-D-glucan 4-glucanohydrolase) gene.";			
RL	J. Bacteriol. 172:3837-3841(1990).			
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLYCOSIDIC LINKAGES IN BETA-D-GLUCANS CONTAINING 1,3- AND 1,4-BONDS.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: M33676; AAA24896.1; -			
DR	PIR: A44507; A44507.			
DR	HSSP; P23904; IAKK.			
DR	InterPro: IPR000757; -			
DR	Pfam: PF00722; Glyco_hydro_16; 1.			
DR	PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.			
KW	Hydrolase; Glycosidase; Signal; Repeat.			
FT	SIGNAL	1	27	
FT	CHAIN	28	349	
FT	ACT_SITE	79	79	BETA-GLUCANASE.
FT	ACT_SITE	83	83	NUCLEOPHILE (BY SIMILARITY).
FT	DOMAIN	271	307	PROTON DONOR (BY SIMILARITY).
FT	REPEAT	271	277	5 X 7 AA TANDEM REPEATS OF P-X-S-S-S-X.
FT	REPEAT	278	284	
FT	REPEAT	285	291	
FT	REPEAT	292	298	
FT	REPEAT	301	307	
SEQ	SEQUENCE	349 AA;	37737 MW;	16DCA4F5BDEFC578A CRC64;

Query Match 100.0%; Score 349; DB 1; Length 349;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKKTVKSGALVAAALTTTVKSAKDSGAEVLTLEVOYKGFAPKMAAAGTSS 60
 DB 1 MNKKTVKSGALVAAALTTTVKSAKDSGAEVLTLEVOYKGFAPKMAAAGTSS 60
 QY 61 SMFLYONGSEIADGPRVVEVDIEVLGNKPSFOSNITTTGAKAQKTSEKHNVSPADQA 120
 DB 61 SMFLYONGSEIADGPRVVEVDIEVLGNKPSFOSNITTTGAKAQKTSEKHNVSPADQA 120
 QY 121 FHYYGLEMTNRYRYMYVDGDEVRKTEGGVSNLTGTGLFPLNLSAAWVGQFDESKL 180
 DB 121 FHYYGLEMTNRYRYMYVDGDEVRKTEGGVSNLTGTGLFPLNLSAAWVGQFDESKL 180
 QY 181 PLFOFIMVWVYKTPGCGSGSDFLTMDTNDNPTDFGSRMGKDMTFDGNRDLTCKNI 240
 DB 181 PLFOFIMVWVYKTPGCGSGSDFLTMDTNDNPTDFGSRMGKDMTFDGNRDLTCKNI 240
 QY 241 YSRDGMILIALTRKGSFNGVPRDEPAPQSSSSAPASSSSVPAASSSSAFV 300
 DB 241 YSRDGMILIALTRKGSFNGVPRDEPAPQSSSSAPASSSSVPAASSSSAFV 300
 QY 301 PPSSSSTNTAIGHKRTTPAVAKEHRLVNAKAKVNPNGHKRRVNEEH 349
 DB 301 PPSSSSTNTAIGHKRTTPAVAKEHRLVNAKAKVNPNGHKRRVNEEH 349

RESULT 2
 HAIR_DROME STANDARD; PRT; 337 AA.
 AC P14003; Q9VSN8;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HAIRY PROTEIN.
 GN H OR CG6494.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Anthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCB1_taxid=227;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=90059896; PubMed=2479541;
 RA Rushlow C.A., Hogan A., Pierchin S.M., Howe K.M., Iardelli M.,
 RA Ish-Horowicz D.;
 RT "The Drosophila hairy protein acts in both segmentation and bristle
 RT patterning and shows homology to N-myc.";
 RL EMBO J. 8:3095-3103(1989).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abail J.F., Adganyi A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Moberly C., Morris J., Mostreft A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venler E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP WRPW MOTIF.
 RX MEDLINE=95094252; PubMed=8001118;
 RA Paroush Z., Finley R.L., Jr., Kidd T., Wainwright S.M., Ingham P.W.,
 RA Brent R., Ish-Horowicz D.;
 RT "Groucho is required for Drosophila neurogenesis, segmentation, and
 RT sex determination and interacts directly with hairy-related bHLH
 RT proteins.";
 RL Cell 79:805-815(1994).
 CC -1- FUNCTION: PAIR-RULE PROTEIN THAT REGULATES EMBRYONIC SEGMENTATION
 CC AND ADULT BRISTLE PATTERNING. TRANSCRIPTIONAL REPRESSOR OF GENES
 CC THAT REQUIRE A BHLH PROTEIN FOR THEIR TRANSCRIPTION (EG. THE FUSHI
 CC TARAZU GENE).
 CC -1- SUBUNIT: TRANSCRIPTION REPRESSION REQUIRES FORMATION OF A COMPLEX
 CC WITH A CO-REPRESSOR PROTEIN (GROUCHO).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- DOMAIN: HAS A PARTICULAR TYPE OF BASIC DOMAIN (PRESENCE OF A
 CC HELIX-INTERRUPTING PROLINE) THAT BINDS TO THE N-BOX (CACNAG),
 CC RATHER THAN THE CANONICAL E-BOX (CANNGG).
 CC -1- DOMAIN: THE CARBOXY-TERMINAL WRPW MOTIF IS A TRANSCRIPTIONAL
 CC REPRESSION DOMAIN NECESSARY FOR THE INTERACTION WITH GROUCHO. A
 CC TRANSCRIPTIONAL CO-REPRESSOR RECRUITED TO SPECIFIC TARGET DNA BY
 CC HAIRY-RELATED PROTEINS.
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. "HAIRY" SUBFAMILY.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X15904; CAA34018.1; -
 CC EMBL: X15905; CAA34019.1; -
 CC EMBL: AEO03554; AAF50378.1; -
 CC PIR: S06956; S06956.
 CC TRANSFAC: T00345; -
 CC Flybase: FBgn0001168; h.
 CC InterPro: IPR001092; -
 CC InterPro: IPR003015; -
 CC Pfam: PF00010; HLH; 1.
 CC PROSITE: PS00038; HELIX_LOOP_HELIX; 1
 CC Nuclear protein; developmental protein; pair-rule protein;
 CC DNA-binding; transcription regulation; repressor; polymorphism.
 CC -----
 CC FT DNA-BINDING 32 44 BASIC DOMAIN.
 CC FT DOMAIN 45 89 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 CC FT DOMAIN 149 157 GIN-RICH.
 CC FT DOMAIN 222 237 GIN-RICH.
 CC FT DOMAIN 241 250 POLY-ALA.
 CC FT VARIANT 292 292 S -> P.
 CC FT DOMAIN 334 337 WRPW MOTIF (REQUIRED FOR ACTIVITY).
 CC SEQUENCE 337 AA; 3695 MW; 6D2ECAF7F2D56C0B CRC64;

Query Match 2.6%; Score 9; DB 1; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 LAVAAAAA 20
DB 238 LAVAAAAA 246

RESULT 3

MMSA_BOVIN STANDARD; PRT; 537 AA.
ID MMSA_BOVIN STANDARD; PRT; 537 AA.
AC 007536;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASE [ACYLATING], MITOCHONDRIAL
DE PRECURSOR (EC 1.2.1.27) (MMSDH).
GN MMSDH.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OX Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX TISSUE=Liver.
RA MEDLINE=93293905; PubMed=8514806;
RA Delchante I., Berthiaume L., Pesceckis S.M., Patton W.F., Resh M.D.;
RT "Novel use of an Iodo-myristyl-CoA analog identifies a semialdehyde
RT dehydrogenase in bovine liver."
RT J. Biol. Chem. 268:13738-13747(1993).
CC -1- FUNCTION: PLAYS A ROLE IN VALINE AND PYRIMIDINE METABOLISM. BINDS
CC FATTY ACYL-CoA.
CC -1- CATALYTIC ACTIVITY: 2-METHYL-3-OXOPROPANATE + COA + NAD(+) =
CC PROPANOL-2-COA + CO(2) + NADH.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- PPM: THE N-TERMINUS IS BLOCKED.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L08643; AAA30650.1; -
DR PIR: A46600; A46600.
DR HSP: P56533; IAS.
DR InterPro: IPR002086; -
DR Pfam: PF00171; aldehyd; 1.
DR PROSITE: PS00070; ALDEHYDE DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE DEHYDR_GLU; FALSE NEG.
KW Oxidoreductase; NAD; Transit peptide; Mitochondrion.
FT TRANSIT 1 34
FT CHAIN 35 537
FT NP_BIND 263 268
FT ACT_SITE 319 319
FT CONFLICT 126 127
FT SEQUENCE 537 AA; 58062 MW; 69DF39506E62F9C0 CRC64;
ML -> TD (IN AA SEQUENCE).
BY SIMILARITY.
Query Match 2.6%; Score 9; DB 1; Length 537;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 13 AVAAAAAL 21

DB 5 AVAAAAAL 13

Query Match 2.6%; Score 9; DB 1; Length 537;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 LAVAAAAA 20
DB 238 LAVAAAAA 246

RESULT 4

KPCM_MOUSE STANDARD; PRT; 918 AA.
ID KPCM_MOUSE STANDARD; PRT; 918 AA.
AC 062101;
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE PROTEIN KINASE C, MU TYPE (EC 2.7.1.1) (NPKC-MU) (PROTEIN KINASE D).
DE PKCKM OR PKCM OR PKD.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BALB/C; TISSUE=Lung;
RX MEDLINE=94359973; PubMed=8078925;
RX Valverde A.M., Smet-Smit J., Van Lint J., Rozenburg E.;
RT "Molecular cloning and characterization of protein kinase D: a target
RT for diacylglycerol and phorbol esters with a distinctive catalytic
RT domain."
RT Proc. Natl. Acad. Sci. U.S.A. 91:8572-8576(1994).
CC -1- FUNCTION: THIS IS CALCIUM-INDEPENDENT, PHOSPHOLIPID-DEPENDENT,
CC SERINE- AND THREONINE-SPECIFIC ENZYME.
CC -1- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.
CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC PKC SUBFAMILY.
CC -----
DR EMBL: Z34524; CAA84283.1; -
DR HSP: P28867; IPTQ.
DR MGD: MGI:99879; PKcm.
DR InterPro: IPR000719; -
DR InterPro: IPR001849; -
DR InterPro: IPR002219; -
DR InterPro: IPR002290; -
DR Pfam: PF00069; PKkinase; 1.
DR Pfam: PF00130; DAG_PE-bind; 2.
DR Pfam: PF00169; PH; 1.
DR PRINTS: PR00008; DAGPEDOMAIN.
DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 2.
DR PROSITE: PS00081; DAG_PE_BIND_DOM_2; 2.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase; Serine/threonine-protein kinase;
KW Phorbol-ester binding; Zinc; Repeat.
FT DOMAIN 145 194
FT DOMAIN 277 326
FT DOMAIN 428 547
FT DOMAIN 589 845
FT DOMAIN 16 26
FT DOMAIN 198 201
FT NP_BIND 595 603
FT BINDING 618 618
FT ACT_SITE 712 712
FT BINDING 618 618
FT ACT_SITE 712 712
BY SIMILARITY.
Query Match 2.6%; Score 9; DB 1; Length 537;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 13 AVAAAAAL 21

SO SEQUENCE 918 AA; 102067 MW; 234486180521BDDBA CRC64;

Query Match

Best Local Similarity 2.6%; Score 9; DB 1; Length 918;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 AVAAAAAAL 21
19 AVAAAAAAL 27

RESULT 5

ID N153_HUMAN STANDARD; PRT; 1475 AA.

AC P49790;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE NUCLEAR PORE COMPLEX PROTEIN NUP153 (NUCLEOPORIN NUP153) (153 KDA NUCLEOPORIN).

GN NUP153.

OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP MEDLINE=94154002; PubMed=8110839;

RA McMorris I., Bastos R., Horton H., Burke B.;

RT "Sequence analysis of a cDNA encoding a human nuclear pore complex protein, hnup153."

RL Biochim. Biophys. Acta 1217:219-223(1994).

CC -1- FUNCTION: POSSIBLE DNA-BINDING SUBUNIT OF THE NUCLEAR PORE COMPLEX (NPC). THE REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN

ANCHORING COMPONENTS OF THE PORE COMPLEX TO THE PORE MEMBRANE.

CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX. LOCATED TO THE

TERMINAL RING STRUCTURE OF THE NUCLEOPLASMIC CAGE.

CC -1- DOMAIN: CONTAINS X-F-X-F-G REPEATS.

CC -1- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS

SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST

NUP1, NSP1, POM 121 AND MAMMALIAN P62.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC -----

DR EMBL: Z25535; CAA80982.1; -

DR MIM: 603948; -

DR InterPro: IPR001876; -

DR Pfam: PF00641; zf-RanBP; 4.

KW Nuclear protein; Transport; Repeat; Zinc-finger; DNA-binding.

FT DOMAIN 4 14

FT ZN_FING 662 681

FT ZN_FING 726 745

FT ZN_FING 797 816

FT ZN_FING 855 874

SO SEQUENCE 1475 AA; 153889 MW; 3CB415A6909DF80E CRC64;

Query Match 2.6%; Score 9; DB 1; Length 1475;

Best Local Similarity 100.0%; Pred. No. 3.8;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 282 SSVPASSSS 290

DB 823 SSVPASSSS 831

RESULT 6

ID ANP3_PSEAM STANDARD; PRT; 37 AA.

AC P02733;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last annotation update)

DE ANTIFREEZE PEPTIDE 3.

OS Pseudopleuronectes americanus (Winter flounder).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percormorpha; Pleuronectiformes;

OC Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.

OX NCBI_TaxID=8265;

RN [1]

RP SEQUENCE

RX MEDLINE=78060969; PubMed=588591;

RA Davies A.L., Lin Y.;

RT "Structure of a peptide antifreeze and mechanism of adsorption to

ice."

RL Biochim. Biophys. Acta 495:388-392(1997).

CC -1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.

CC -1- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE I AFP ARE

ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.

DR PIR: A03192; PFEL3W.

DR HSSP: P04002; 1WFA.

DR InterPro: IPR00104; -

DR PRINTS: PR00308; ANTIFREEZE1.

KW Antifreeze protein; Repeat; Multigene family.

SO SEQUENCE 37 AA; 3144 MW; 46AA951A962DECA9 CRC64;

OY 15 AVAAAAAL 22

DB 6 AVAAAAAL 13

RESULT 7

ID ANP4_PSEAM STANDARD; PRT; 82 AA.

AC P04002;

DT 23-OCT-1986 (Rel. 02, Created)

DT 23-OCT-1986 (Rel. 02, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE ANTIFREEZE PROTEIN A/B PRECURSOR.

OS Pseudopleuronectes americanus (Winter flounder).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percormorpha; Pleuronectiformes;

OC Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.

OX NCBI_TaxID=8265;

RN [1]

RP SEQUENCE FROM N.A. (PROTEIN A).

RX MEDLINE=82197490; PubMed=6952188;

RA Davies P.L., Roach A.H., Hew C.-L.;

RT "DNA sequence coding for an antifreeze protein precursor from winter

flounder."

RT Proc. Natl. Acad. Sci. U.S.A. 79:335-339(1982).

SO SEQUENCE FROM N.A. (PROTEIN A).

RX MEDLINE=88259236; PubMed=3133486;

RA Scott G.K., Davies P.L., Kao M.H., Fletcher G.L.;

RT "Differential amplification of antifreeze protein genes in the

pleuronectinae."

SO J. Mol. Evol. 27:29-35(1988).

RN [3]

RP SEQUENCE FROM N.A. (PROTEIN B).

RX MEDLINE=84264559; PubMed=6086629;

RA Davies P.L., Hough C., Scott G.K., Ng N., White B.N., Hew C.-L.;
 RT "Antifreeze protein genes of the winter flounder.";
 RL J. Biol. Chem. 259:9241-9247(1984).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92209995; PubMed=1555765;
 RA Davies P.L.;
 RT "Conservation of antifreeze protein-encoding genes in tandem
 repeats.";
 RL Gene 112:163-170(1992).
 RN [5]
 RP 3D-STRUCTURE MODELING OF 45-81.
 RX MEDLINE=92148833; PubMed=1738160;
 RA Chou K.-C.;
 RT "Energy-optimized structure of antifreeze protein and its binding
 mechanism.";
 RL J. Mol. Biol. 223:509-517(1992).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 45-81.
 RX MEDLINE=95281060; PubMed=7760940;
 RA Sichert F., Yang D.S.C.;
 RT "Ice-binding structure and mechanism of an antifreeze protein from
 winter flounder.";
 RL Nature 375:427-431(1995).
 CC -I- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD-FREEZING POINT.
 CC -I- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF PROTEIN A.
 CC -I- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE 1 AFP ARE
 CC ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC -----
 DR EMBL; L00138; AAB59964.1; -
 DR EMBL; L29178; AAB59964.1; JOINED.
 DR EMBL; M62414; AAA49469.1; -
 DR EMBL; X07506; CAA30389.1; -
 DR EMBL; M62416; AAA49471.1; -
 DR EMBL; M62417; AAA49472.1; -
 DR PIR; A03194; FDFLAW.
 DR PIR; A05161; A05161.
 DR PIR; S02326; S02326.
 DR PIR; JS0704; JS0704.
 DR PDB; 1ATE; 15-OCT-94.
 DR PDB; 1WFA; 03-JUN-95.
 DR PDB; 1WFB; 03-JUN-95.
 DR InterPro: IPR000104; -
 DR PRINTS: PR00308; ANTIFREEZE1.
 KW Antifreeze protein; Repeat; Multigene family; Signal; 3D-structure.
 FT SIGNAL 1 21
 FT PROPEP 22 44
 FT
 FT CHAIN 45 82 REMOVED BY A DIPEPTIDYLPEPTIDASE
 FT VARIANT 36 36 (PROBABLY).
 FT VARIANT 70 70 A -> V.
 FT CONFLICT 24 24 A -> D (IN PROTEIN B).
 FT HELIX 46 80 S -> R (IN REF. 2).
 SO SEQUENCE 82 AA; 7711 MW; C2AE7B74C0D46CC1 CRC64;

Query Match 2.38; Score 8; DB 1; Length 82;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 AAAAALT 22
 DB 50 AAAAALT 57

RESULT 8
 ZEST_DROMA STANDARD; PRT; 268 AA.
 ID ZEST_DROMA
 AC Q24597; Q24598; Q24599; Q24600; Q24601; Q24602; Q24603; Q24604;
 AC Q27387;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE REGULATORY PROTEIN ZESTE (FRAGMENT).
 GN 2.
 OS Drosophila mauritiana (Fruit fly),
 OS Drosophila sechellia (Fruit fly), and
 OS Drosophila simulans (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7226, 7238, 7240;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VARIOUS STRAINS;
 RX MEDLINE=93360802; PubMed=8355601;
 RA Hey J., Kilman R.M.;
 RT "Population genetics and phylogenetics of DNA sequence variation at
 RT multiple loci within the Drosophila melanogaster species complex.";
 RL Mol. Biol. Evol. 10:804-822(1993).
 CC -I- FUNCTION: INVOLVED IN TRANSECTON PHENOMENA (= SYNOPSIS-DEPENDENT
 CC GENE EXPRESSION), WHERE THE SYNAPTIC PAIRING OF CHROMOSOMES
 CC CARRYING GENES WITH WHICH ZESTE INTERACTS INFLUENCES THE
 CC EXPRESSION OF THESE GENES. ZESTE BINDS TO DNA AND STIMULATES
 CC TRANSCRIPTION FROM A NEARBY PROMOTER (BY SIMILARITY).
 CC -I- SUBUNIT: SELF-ASSOCIATE FORMING COMPLEXES OF SEVERAL HUNDRED
 CC MONOMERS (BY SIMILARITY).
 CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC -----
 DR EMBL; L13049; AAA29032.1; -
 DR EMBL; L13050; AAA29033.1; -
 DR EMBL; L13051; AAA29034.1; -
 DR EMBL; L13052; AAA29035.1; -
 DR EMBL; L13053; AAA29036.1; -
 DR EMBL; L13054; AAA29037.1; -
 DR EMBL; L13055; AAA29038.1; -
 DR EMBL; L13056; AAA29039.1; -
 DR EMBL; L13057; AAA29040.1; -
 DR EMBL; L13058; AAA29041.1; -
 DR EMBL; L13059; AAA29042.1; -
 DR EMBL; L13060; AAA29043.1; -
 DR EMBL; L13061; AAA29044.1; -
 DR EMBL; L13062; AAA29045.1; -
 DR EMBL; L13063; AAA29046.1; -
 DR EMBL; L13064; AAA29047.1; -
 DR EMBL; L13065; AAA29048.1; -
 DR EMBL; L13066; AAA29049.1; -
 DR FlyBase: FBgn0012518; Dmanvz.
 DR FlyBase: FBgn0012801; Dsecvz.
 DR FlyBase: FBgn0012904; Dsinvz.
 KW DNA-binding; Transcription regulation; Nuclear protein.
 FT DNA_BIND 1 1
 FT NON_TER 1 1
 FT DOMAIN 97 268
 FT
 FT VARIANT 105 106 SPECIFIC, WITH ZESTE LOCUS.
 FT NON_TER 268 268 GIN/ALA-RICH (OPA-REPEAT INVOLVED IN
 FT TRANSCRIPTIONAL ACTIVATION OR REPRESSION
 FT AT DIFFERENT TARGET LOCI) (POTENTIAL).
 FT MISSING (IN STRAIN SI-K2).
 SO SEQUENCE 268 AA; 30353 MW; F592F8FA4F2DB47B CRC64;

Query Match 2.3%; Score 8; DB 1; Length 268;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 AVAAAAA 20
 Db 192 AVAAAAA 199

RESULT 9
 CYSD_RHIME STANDARD; PRT; 317 AA.
 ID CYSD_RHIME
 AC P56892;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SULFATE ADENYLYLTRANSFERASE SUBUNIT 2 (EC 2.7.7.4) (SULFATE ADENYLYLTRANSFERASE) (SAT) (ATP-SULFURYLASE SMALL SUBUNIT).
 GN CYSD.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 NC NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=99395034; PubMed=10464198;
 RA Abola A.P., Willis M.G., Wang R.C., Long S.R.;
 RT "Reduction of adenosine-5'-phosphosulfate instead of 3'-phosphodensinate-5'-phosphosulfate in cysteine biosynthesis by Rhizobium meliloti and other members of the family Rhizobiaceae.";
 RL J. Bacteriol. 181:5280-5287(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA Galtier F., Capela D., Hubler-Barloy F., Gattus M., Batut J., Boistard P., Guzy J., Kahn D., Thebaud P., Goffeau A., Purnelle B., Pohl T., Bothe G., Schneider S., Portetelle D., Vandenbol M., Puchler A., Becker A., Weidner S.;
 RA Submitted (MAY-2000) to the SWISS-PROT data bank;
 RL
 CC -1- CATALYTIC ACTIVITY: ATP + SULFATE -> PYROPHOSPHATE + ADENYLYLSULFATE.
 CC -1- PATHWAY: FIRST STEP IN THE SULFATE ACTIVATION PATHWAY. THIS REACTION OCCURS EARLY IN THE REDUCTIVE BRANCH OF THE CYSTEINE BIOSYNTHETIC PATHWAY.
 CC -1- SUBUNIT: HETERODIMER COMPOSED OF CYSD, THE SMALLER SUBUNIT, AND -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY. CYSD SUBFAMILY.
 CC -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY. CYSD SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF158023; AAD55760.1; -
 CC DR InterPro: IPR002500; -
 CC DR Pfam: PF01507; PAPS_reduct; 1.
 CC DR Cysteine biosynthesis; Transferase; Nucleotidyltransferase.
 CC KW CONFLICT 260 260 R -> A (IN REF. 1).
 CC FT SEQUENCE 317 AA; 36429 MW; 093FE274D22841CB CRC64;

Query Match 2.3%; Score 8; DB 1; Length 317;
 Best Local Similarity 100.0%; Pred. No. 8.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 243 RDGMILIA 250
 Db 234 RDGMILIA 241

RESULT 10
 YE80_MYCTU STANDARD; PRT; 317 AA.
 ID YE80_MYCTU
 AC P71761; 053171;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHETICAL 34.3 KDA PROTEIN RV1480.
 GN RV1480 OR MTW007.27 OR MYCY277.01.
 GN
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NC NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F., Davies R., Devlin K., Jagels K., Felwell T., Gentles S., Hamlin N., Holtroyd S., Hornsby T., Jagers K., Kirogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skellon S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";
 RT Nature 393:537-544(1998).
 RL Nature 393:537-544(1998).
 CC -1- SIMILARITY: STRONG, TO M.AVIUM MAV169.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AL021184; CA116010.1; -
 CC DR Tuberculist: RV1480; -
 CC DR Hypothetical protein.
 CC KW SEQUENCE 317 AA; 34333 MW; 3D1D060FF5E5CE56 CRC64;

Query Match 2.3%; Score 8; DB 1; Length 317;
 Best Local Similarity 100.0%; Pred. No. 8.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 LAVAAAAA 19
 Db 120 LAVAAAAA 127

RESULT 11
 CHEB_PSEAE STANDARD; PRT; 368 AA.
 ID CHEB_PSEAE
 AC O87125;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE PROTEIN-GLUTAMATE METHYLESTERASE (PC 3.1.1.61).
 GN CHEB OR PA1459.
 GN
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OC NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PAO1;
 RX MEDLINE=99161288; PubMed=10052136;
 RA Kato J., Nakamura T., Kuroda A., Ohtake H.;

"Cloning and characterization of chemotaxis genes in *Pseudomonas aeruginosa*.";
 Biosci. Biotechnol. Biochem. 63:155-161(1999).
 [2]
 SEQUENCE FROM N.A.
 STRAIN=PA01;
 MEDLINE=20437337; PubMed=10984043;
 Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P., Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M., Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T., Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
 "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.";
 Nature 406:959-964(2000).
 CC -!- FUNCTION: INVOLVED IN THE MODULATION OF THE CHEMOTAXIS SYSTEM;
 CC CATALYZES THE DEMETHYLATION OF SPECIFIC METHYLGLUTAMATE RESIDUES
 CC INTRODUCED INTO THE CHEMORECEPTORS (METHYL-ACCEPTING CHEMOTAXIS
 CC PROTEINS) BY CHER (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: PROTEIN L-GLUTAMATE O-METHYL ESTER + H(2)O =
 CC PROTEIN L-GLUTAMATE + METHANOL.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- DOMAIN: THE N-TERMINAL REGULATORY DOMAIN INHIBITS THE ACTIVITY OF
 CC THE C-TERMINAL EFFECTOR DOMAIN.
 CC -!- PTM: PHOSPHORYLATED BY CHEA. PHOSPHORYLATION SUPPRESSES THE
 CC INHIBITORY ACTIVITY OF THE N-TERMINAL DOMAIN (BY SIMILARITY).
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE RESPONSE
 CC REGULATORY FAMILY.
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE CHEB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AB012767; BAA33550.1; -;
 DR EMBL; AE004575; AAG04848.1; -;
 DR HSSP; P04042; ICHD.
 DR InterPro; IPR000673; -;
 DR InterPro; IPR001789; -;
 DR Pfam; PF01339; CheB_methyltest; 1.
 DR Pfam; PF00072; response_reg; 1.
 KW Hydrolase; Chemotaxis; Sensory transduction; Phosphorylation.
 FT DOMAIN 1 133 RESPONSE REGULATORY DOMAIN.
 FT DOMAIN 134 177 LINKER.
 FT DOMAIN 178 368 PROTEIN-GLUTAMATE METHYLESTERASE.
 FT MOD_RES 55 55 PHOSPHORYLATION (BY SIMILARITY).
 FT ACT_SITE 192 192 BY SIMILARITY.
 FT ACT_SITE 219 219 BY SIMILARITY.
 FT ACT_SITE 312 312 BY SIMILARITY.
 FT CONFLICT 34 34 G -> A (IN REF. 1).
 SQ SEQUENCE -368 AA; 39004 MW; FE4801DC220C613B CRC64;
 Query Match 2.3%; Score 8; DB 1; Length 368;
 Best Local Similarity 100.0%; Pred. No. 9.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 277 APASSSSV 284
 Db 151 APASSSSV 158
 RESULT 12
 SOX1_HUMAN
 ID SOX1_HUMAN STANDARD; PRT; 387 AA.
 AC O00570;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

15-JUL-1998 (Rel. 36, Last annotation update)
 DE SOX-1 PROTEIN.
 GN SOX1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98051911; PubMed=9337405;
 RA Malas S., Duthie S.M., Mohri F., Lovell-Badge R., Episkopou V.;
 "Cloning and mapping of the human SOX1: a highly conserved gene
 RT expressed in the developing brain.";
 RL Mamm. Genome 8:866-868(1997).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -!- TISSUE SPECIFICITY: MAINLY IN THE DEVELOPING CENTRAL NERVOUS
 CC SYSTEM.
 CC -!- SIMILARITY: CONTAINS 1 HMG BOX.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; Y13436; CAA73847.1; -;
 DR HSSP; Q05066; IHRZ.
 DR MIM; 602148; -;
 DR InterPro; IPR000910; -;
 DR Pfam; PF00505; HMG_box; 1.
 KW DNA-binding; Nuclear protein.
 FT DOMAIN 27 43
 FT DNA_BIND 51 119 HMG_BOX.
 FT DOMAIN 145 150 POLY-GLY.
 FT DOMAIN 197 204 POLY-GLY.
 FT DOMAIN 280 287 POLY-ALA.
 FT DOMAIN 292 302 POLY-ALA.
 FT DOMAIN 353 360 POLY-ALA.
 SQ SEQUENCE 387 AA; 38855 MW; 8D43A38002CE3494 CRC64;
 Query Match 2.3%; Score 8; DB 1; Length 387;
 Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 13 AVAAAAAA 20
 Db 292 AVAAAAAA 299
 RESULT 13
 HME1_HUMAN
 ID HME1_HUMAN STANDARD; PRT; 391 AA.
 AC Q05925;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HOMEBOX PROTEIN ENGRAILED-1 (HU-EN-1).
 GN EN1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93185339; PubMed=1363401;
 RA Logan C., Hanks M.C., Noble-Topham S., Nallainathan D.,
 RA Provart N.J., Joyner A.L.;
 RT "Cloning and sequence comparison of the mouse, human, and chicken
 RT engrailed genes reveal potential functional domains and regulatory
 RT regions.";

RL Dev. Genet. 13:345-358(1992).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- SIMILARITY: BELONGS TO THE ENGRAILED FAMILY OF HOMEOBOX PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L12698; AAA53501.1; -;
 CC EMBL; L12699; AAA53502.1; -;
 CC HSSP; P02836; LENH.
 CC TRANSFAC; T02015; -;
 CC MTM; L31290; -;
 CC InterPro; IPR000747; -;
 CC InterPro; IPR001356; -;
 CC Pfam; PF00046; homeobox; 1.
 CC PRINTS; PR00024; HOMEOBOX.
 CC PROSITE; PS00026; ENGRAILED.
 CC PROSITE; PS00027; HOMEOBOX_1; 1.
 CC PROSITE; PS00071; HOMEOBOX_2; 1.
 CC PROSITE; PS00033; ENGRAILED; 1.
 CC Homeobox; DNA-binding; Developmental protein; Nuclear protein.
 KW Homeobox; DNA-binding; ENGRAILED; 1.
 FT DOMAIN 199 218 POLY-ALA.
 FT DOMAIN 224 231 POLY-GLY.
 FT DNA_BIND 302 361 HOMEOBOX.
 FT SEQUENCE 391 AA; 40044 MW; 30DA29BF0995956F CRC64;
 SQ
 Query Match 2.3%; Score 8; DB 1; Length 391;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 13 AVAAAAAA 20
 Db 208 AVAAAAAA 215
 RESULT 14
 SOX1_MOUSE STANDARD; PRT; 391 AA.
 AC P53783;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE SOX-1 PROTEIN.
 GN SOX1 OR SOX-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RX MEDLINE=96189340; PubMed=8625802;
 RA Collignon J., Sockanathan S., Hacker A., Cohen-Tannoudji M.,
 RA Norris D., Rastan S., Stevanovic M., Goodfellow P.N.,
 RA Lovell-Badge R.;
 RT "A comparison of the properties of Sox-3 with Sry and two related
 RT genes, Sox-1 and Sox-2.";
 RL Development 122:509-520(1996).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -!- TISSUE SPECIFICITY: MAINLY IN THE DEVELOPING CENTRAL NERVOUS
 CC SYSTEM. EXPRESSED IN DEVELOPING UROGENITAL RIDGE.
 CC -!- SIMILARITY: CONTAINS 1 HMG BOX.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X94126; CAA63846.1; -;
 CC HSSP; Q05066; LHR2.
 CC MGD; MGI:98357; Sox1.
 CC InterPro; IPR000910; -;
 CC Pfam; PF00505; HMG-box; 1.
 CC DNA-binding; Nuclear protein.
 KW Homeobox; DNA-binding; ENGRAILED; 1.
 FT DOMAIN 30 43 POLY-GLY.
 FT DNA_BIND 51 119 HMG BOX.
 FT DOMAIN 145 150 POLY-GLY.
 FT DOMAIN 197 204 POLY-ALA.
 FT DOMAIN 280 288 POLY-ALA.
 FT DOMAIN 296 306 POLY-ALA.
 FT DOMAIN 357 364 POLY-ALA.
 FT SEQUENCE 391 AA; 39237 MW; 9F81ED667F947C05 CRC64;
 SQ
 Query Match 2.3%; Score 8; DB 1; Length 391;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 13 AVAAAAAA 20
 Db 296 AVAAAAAA 303
 RESULT 15
 HMEL_MOUSE STANDARD; PRT; 401 AA.
 ID HMEL_MOUSE STANDARD; PRT; 401 AA.
 AC P09065;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HOMEOBOX PROTEIN ENGRAILED-1 (MO-EN-1).
 GN EN1 OR EN-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93185339; PubMed=1363401;
 RA Logan C., Hanks M.C., Noble-Topham S., Nallainathan D.,
 RA Provart N.J., Joyner A.L.;
 RT "Cloning and sequence comparison of the mouse, human, and chicken
 RT engrailed genes reveal potential functional domains and regulatory
 RT regions.";
 RL Dev. Genet. 13:345-358(1992).
 RN [2]
 RP SEQUENCE OF 278-401 FROM N.A.
 RX MEDLINE=88112776; PubMed=2892757;
 RA Joyner A.L., Martin G.R.;
 RT "En-1 and En-2, two mouse genes with sequence homology to the
 RT Drosophila engrailed gene: expression during embryogenesis.";
 RL Genes Dev. 1:29-38(1987).
 RN [3]
 RP SEQUENCE OF 298-401 FROM N.A.
 RX MEDLINE=86079501; PubMed=2416459;
 RA Joyner A.L., Kornberg T., Coleman K.G., Cox D.R., Martin G.R.;
 RT "Expression during embryogenesis of a mouse gene with sequence
 RT homology to the Drosophila engrailed gene.";
 RL Cell 43:29-37(1985).
 RN [4]
 RP SEQUENCE OF 321-380 FROM N.A.
 RX MEDLINE=91099509; PubMed=1980115;
 RA Holland P.W.H., Williams N.A.;
 RT "Conservation of engrailed-like homeobox sequences during vertebrate
 RT evolution.";
 RL FEBS Lett. 277:250-252(1990).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC -!- SIMILARITY: BELONGS TO THE ENGRAILED FAMILY OF HOMEBOX PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L12703; AAA03660.2; -
CC EMBL; Y00201; CAA68361.1; -
CC PIR; A26629; A26629.
CC PIR; A24778; A24778.
CC PIR; S13009; S13009.
CC PIR; A48423; A48423.
CC TRANSFAC; T02016; -
CC MGD; MGI:95389; Enl.
CC InterPro; IPR000747; -
CC InterPro; IPR001356; -
CC Pfam; PF00046; homeobox; 1.
CC PRINTS; PR00024; HOMEBOX.
CC PRINTS; PR00026; ENGRAILED.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS00071; HOMEBOX_2; 1.
CC PROSITE; PS00033; ENGRAILED; 1.
CC Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DOMAIN 52 87 PRO-RICH.
FT DOMAIN 73 87 POLY-PRO.
FT DOMAIN 207 228 POLY-ALA.
FT DNA_BIND 312 371 HOMEBOX.
SQ SEQUENCE 401 AA; 40950 MW; 1F90210950152FAE CRC64;

Query Match 2.3%; Score 8; DB 1; Length 401;
Best Local Similarity 100.0%; Pred No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 AVAAAAAA 20
|
Db 220 AVAAAAAA 227

Search completed: August 23, 2001, 11:18:30
Job time: 197 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2001, 11:19:26 ; Search time 50.81 Seconds
(without alignments)
908.767 Million cell updates/sec

Title: US-09-654-652A-3
Perfect score: 349
Sequence: 1 MNIKTAVKSALAVAAAAA.....AKGAKVNPNGHKRYRVNFHEH 349

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 425026 seqs, 132305027 residues

Word size : 0

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_16.*
1: sp-archaea.*
2: sp-bacteria.*
3: sp-fungi.*
4: sp-human.*
5: sp-invertebrate.*
6: sp-mammal.*
7: sp-mhc.*
8: sp-organelle.*
9: sp-phage.*
10: sp-plant.*
11: sp-rodent.*
12: sp-unclassified.*
13: sp-vertebrate.*
14: sp-virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	3.2	539	10 Q9SR71	Q9sr71 arabidopsis
2	10	2.9	250	10 Q9LJ08	Q9lj08 oryza sativ
3	10	2.9	348	10 Q96517	Q96517 arabidopsis
4	10	2.9	348	10 Q80398	Q80398 arabidopsis
5	10	2.9	369	10 Q04322	Q04322 arabidopsis
6	10	2.9	621	2 Q59301	Q59301 cellvibrio
7	9	2.6	87	10 Q9LJ13	Q9lj13 oryza sativ
8	9	2.6	184	1 Q9HQ55	Q9hq55 halobacteri
9	9	2.6	300	2 Q9KWF8	Q9kwf8 kitasatospo
10	9	2.6	321	7 Q9XS19	Q9xs19 papio hamad
11	9	2.6	350	2 Q9KMG4	Q9kmg4 streptomyc
12	9	2.6	439	1 Q9YCK0	Q9yck0 aeropyrum p
13	9	2.6	518	10 Q9PRC3	Q9prc3 oryza sativ
14	9	2.6	776	2 Q86516	Q86516 streptomyc
15	9	2.6	840	5 Q9NB03	Q9nb03 aedes aegyp
16	9	2.6	1068	5 Q9W3X2	Q9w3x2 drosophila
17	9	2.6	1307	5 Q9V702	Q9v702 drosophila
18	9	2.6	3112	5 Q9NKP1	Q9nkp1 leishmania
19	8	2.3	23	4 Q43519	Q43519 homo sapien

20	8	2.3	38	10	Q9S9D9	Q9s9d9 nicotiana t
21	8	2.3	46	7	P79534	P79534 homo sapien
22	8	2.3	46	7	P79536	P79536 homo sapien
23	8	2.3	46	7	P79538	P79538 homo sapien
24	8	2.3	67	5	Q9NNB6	Q9nnb6 leishmania
25	8	2.3	82	13	Q9YH51	Q9yh51 pseudopleur
26	8	2.3	82	13	Q99013	Q99013 pseudopleur
27	8	2.3	101	10	Q9LGG6	Q9lgg6 oryza sativ
28	8	2.3	107	4	Q16846	Q16846 homo sapien
29	8	2.3	111	4	Q15587	Q15587 homo sapien
30	8	2.3	134	4	Q43816	Q43816 homo sapien
31	8	2.3	135	4	Q9NP14	Q9np14 homo sapien
32	8	2.3	137	6	Q29365	Q29365 sus scrofa
33	8	2.3	138	4	P78428	P78428 homo sapien
34	8	2.3	139	6	Q9GLV8	Q9glv8 equus cabal
35	8	2.3	141	7	Q9GJ06	Q9gj06 homo sapien
36	8	2.3	152	4	Q00496	Q00496 homo sapien
37	8	2.3	152	5	Q9U730	Q9u730 drosophila
38	8	2.3	152	5	Q9TVX1	Q9tvx1 drosophila
39	8	2.3	153	5	Q9U731	Q9u731 drosophila
40	8	2.3	153	5	Q9TW37	Q9tw37 drosophila
41	8	2.3	154	5	Q9TW44	Q9tw44 drosophila
42	8	2.3	158	5	O02055	O02055 caenorhabdi
43	8	2.3	175	10	Q9LGB7	Q9lgb7 oryza sativ
44	8	2.3	197	2	Q9X6J7	Q9x6j7 vibrio chol
45	8	2.3	198	10	Q9FTW8	Q9ftw8 oryza sativ

ALIGNMENTS

RESULT 1

Q9SR71 ID Q9SR71 PRELIMINARY; PRT; 539 AA.
AC Q9SR71;
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DE T22K18.10 PROTEIN.
GN T22K18.10.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T22K18 genomic sequence.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC010927; AAF04417.1;
DR InterPro; IPR000166;
DR InterPro; IPR002965;
DR PRINTS; PR01217; PRICHEXTENSN.
SQ SEQUENCE 539 AA; 57710 MW; 72597A2E1B57EE9B CRC64;

Query Match 3.2%; Score 11; DB 10; Length 539;
Best Local Similarity 100.0%; Pred. No. 0.049;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 SSSSAPASSSS 283

Db 117 SSSSAPASSSS 127

RESULT 2

Q9LJ08 ID Q9LJ08 PRELIMINARY; PRT; 250 AA.
AC Q9LJ08;

DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DE ESTS AU067519(C10906).
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
 OC Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0667A10.";
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF001073; BAA89582.1; -;
 DR InterPro; IPR000010; -;
 DR Pfam; PF00031; cystatin; 2.
 DR PROSITE; PS00287; CYSTATIN; UNKNOWN_1.
 DR SMART; SM00043; CY; 1.
 SQ SEQUENCE 250 AA; 27269 MW; 6F2A7D4EB8AF66F CRC64;

Query Match 2.9%; Score 10; DB 10; Length 250;
 Best Local Similarity 100.0%; Pred. No. 0.23;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LAVAAAAAL 21
 DB 24 LAVAAAAAL 33

RESULT 3
 ID Q96517 PRELIMINARY; PRT; 348 AA.
 AC Q96517;
 DT 01-FEB-1997 (TReMBLrel. 02, Created)
 DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE MAP KINASE KINASE ALPHA PROTEIN KINASE (EC 2.7.1.37) (PHOSPHORYLASE B
 DE KINASE KINASE) (GLYCOGEN SYNTHASE A KINASE) (HYDROXYALKYL-PROTEIN
 DE KINASE) (SERINE(THREONINE) PROTEIN KINASE).
 GN MAP2KALPHA.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=97092864; PubMed=8938426;
 RA Jouannic S., Hamal A., Kreis M., Henry Y.;
 RL Plant Physiol. 112:1397-1397(1996).
 DR EMBL; Y07694; CAA68958.1; -;
 DR Mendel; 14316; Arath; 2349; 14316.
 DR InterPro; IPR000719; -;
 DR InterPro; IPR002290; -;
 DR Pfam; PF00069; pkinase; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
 DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
 DR SMART; SM00220; S_TKc; 1.
 KW ATP-binding; Kinase; Transferase.
 SQ SEQUENCE 348 AA; 38329 MW; C18C987C6B7AF5F5 CRC64;

Query Match 2.9%; Score 10; DB 10; Length 348;
 Best Local Similarity 100.0%; Pred. No. 0.3;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 SSSAPASSS 282
 DB 46 SSSAPASSS 55

RESULT 4
 ID O80398 PRELIMINARY; PRT; 348 AA.
 AC O80398;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE MAP KINASE KINASE 5.
 GN MKK5.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Shinozaki K., Ichimura K., Mizoguchi T.;
 RT "Molecular cloning and characterization of cDNAs encoding mitogen-
 RT activated protein kinase in Arabidopsis thaliana.";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=20277480; PubMed=10819329;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
 RT clones.";
 RL DNA Res. 7:131-135(2000).
 CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; AB015316; BAA28831.1; -;
 DR EMBL; AB023045; BAB01714.1; -;
 DR HSP; P24941; ICKP.
 DR Mendel; 31061; Arath; 2349; 31061.
 DR InterPro; IPR000719; -;
 DR InterPro; IPR002290; -;
 DR Pfam; PF00069; pkinase; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
 DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR SMART; SM00220; S_TKc; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 348 AA; 38329 MW; 216C96728F34B5B5 CRC64;

Query Match 2.9%; Score 10; DB 10; Length 348;
 Best Local Similarity 100.0%; Pred. No. 0.3;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 SSSAPASSS 282
 DB 46 SSSAPASSS 55

RESULT 5
 ID O04322 PRELIMINARY; PRT; 369 AA.
 AC O04322;
 DT 01-JUL-1997 (TReMBLrel. 04, Created)
 DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE MYB-RELATED TRANSCRIPTION ACTIVATOR (MYBST1) ISOLOG (MYBST1).
 GN T02004.16.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OX Brassicaceae; Brassicaceae; Arabidopsids.
RN NCBI_TaxID=3702;
RP [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Lin X., Ketchum K.A., Phillips C.A., Brandon R.C.,
RA Fuhrmann J.L., White O., Kerlavage A.R., Adams M.D., Somerville C.R.,
RA Venter J.C.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Nakamura Y.;
RL "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
features of the regions of 4,504,864 bp covered by sixty P1 and TAC
clones.";
RL DNA Res. 7:131-135(2000).
DR EMBL; AC001645; AB63650.1; -
DR EMBL; AB023046; BAB01274.1; -
DR Mendel; 17078; Arath;2095:17078.
DR InterPro; IPR001005; -
DR Pfam; PF00249; myb_DNA-binding; 1.
DR PROSITE; PS0090; MYB.3.1.
SQ SEQUENCE 369 AA; 40043 MW; DCC039507044AC9E CRC64;

Query Match 2.9%; Score 10; DB 10; Length 369;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 SALAVAAAAA 19
Db 52 SALAVAAAAA 61
|||||

RESULT 6
Q59301
ID Q59301 PRELIMINARY; PRT; 621 AA.
AC Q59301;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE ENDO-BETA-1,4-XYLANASE PRECURSOR (EC 3.2.1.8) (ENDO-1,4-BETA-XYLANASE)
DE (1,4-BETA-D-XYLAN XYLANOHYDROLASE).
OS Cellvibrio mixtus.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Cellvibrio.
OX NCBI_TaxID=39650;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=MIXTUS;
RC MEDLINE=96077124; PubMed=7492333;
RA Millward-Sadler S.J., Davidson K., Hazlewood G.P., Black G.W.,
RA Gilbert H.J., Clarke J.H.;
RT "Novel cellulose-binding domains, NodB homologues and conserved
modular architecture in xylanases from the aerobic soil bacteria
Pseudomonas fluorescens subsp. cellulosa and Cellvibrio mixtus.";
RL Biochem. J. 312:39-48(1995).
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC
LINKAGES IN XYLANS.
CC EMBL; Z48926; CAA88762.1; -
DR HSSP; P14768; 1CLX.
DR InterPro; IPR001000; -
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRASE10.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; UNKNOWN_1.

KW Signal; Xylan degradation; Hydrolase; Glycosidase.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 621
SQ SEQUENCE 621 AA; 64929 MW; 830E5B959D2C3AB8 CRC64;

Query Match 2.9%; Score 10; DB 2; Length 621;
Best Local Similarity 100.0%; Pred. No. 0.49;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 274 SSSAPASSSS 283
Db 89 SSSAPASSSS 98
|||||

RESULT 7
Q9LL13
ID Q9LL13 PRELIMINARY; PRT; 87 AA.
AC Q9LL13;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN (OSJNBA0086P08.3 PROTEIN).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=CV. NIPPONBARE;
RC Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0708G02.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=CV. NIPPONBARE;
RC Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:OSJNBA0086P08.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP001539; BAA92926.1; -
DR EMBL; AP002855; BAB17194.1; -
SQ SEQUENCE 87 AA; 9123 MW; F3ACD82BA63824EE CRC64;

Query Match 2.6%; Score 9; DB 10; Length 87;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 LAVAAAAA 20
Db 22 LAVAAAAA 30
|||||

RESULT 8
Q9HQS5
ID Q9HQS5 PRELIMINARY; PRT; 184 AA.
AC Q9HQS5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE VNG1026H.
GN VNG1026H.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,

RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrocna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
RT "Genome sequence of Halobacterium species NRC-1";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL: AE005037; AG19438.1; -.
SQ SEQUENCE 184 AA; 18956 MW; B120BC2158DD9095 CRC64;

Query Match 2.6%; Score 9; DB 1; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LAVAAAAA 20
|||||
Db 137 LAVAAAAA 145

RESULT 9
Q9KWF8 ID Q9KWF8 PRELIMINARY; PRT; 300 AA.
AC Q9KWF8;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE MEVALONATE DIPHOSPHATE DECARBOXYLASE.
GN MDPO.
OS Kitasatospora griseola.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Kitasatospora.
OX NCBI_TaxID=2064;
RN [1]
RP SEQUENCE FROM N.A.
RA Daiiri T.;
RT "Cloning of a gene cluster encoding enzymes responsible for the
RT mevalonate pathway from a terpenoid antibiotic-producing Streptomyces
RT strain";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB037907; BAB07818.1; -.
DR InterPro: IPR001745; -.
DR Pfam: PF00288; GHMP_Kinases; 1.
SQ SEQUENCE 300 AA; 30865 MW; 17D9F9D707781997 CRC64;

Query Match 2.6%; Score 9; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 ALAVAAAA 19
|||||
Db 78 ALAVAAAA 86

RESULT 10
Q9XS19 ID Q9XS19 PRELIMINARY; PRT; 321 AA.
AC Q9XS19;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE MHC CLASS I RELATED PROTEIN PRECURSOR (FRAGMENT).
GN MCI.
OS Papio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Papio.
OX NCBI_TaxID=9557;
RN [1]
RP SEQUENCE FROM N.A.

RA Seo J.W., Ross C., Bontrop R., Hunsmann G., Walter L., Guenther E.;
RT "Major histocompatibility complex-linked MIC genes in rhesus macaques
RT and other primates";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
CC IMMUNE SYSTEM (BY SIMILARITY).
CC -I- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN) (BY SIMILARITY).
CC -I- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL: AJ242442; CAB46518.1; -.
DR InterPro: IPR001039; -.
DR InterPro: IPR003006; -.
DR InterPro: IPR003597; -.
DR Pfam: PF00047; Ig, 1.1.
DR Pfam: PF00129; MHC_I, 1.
DR ProDom: PD000050; -. 1.
DR SMART: SM00407; IgC1; 1.
KW Glycoprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 321
SQ SEQUENCE 321 AA; 36039 MW; 8B2579752CE6B65F CRC64;

Query Match 2.6%; Score 9; DB 7; Length 321;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LAVAAAAA 20
|||||
Db 291 LAVAAAAA 299

RESULT 11
Q9KWG4 ID Q9KWG4 PRELIMINARY; PRT; 350 AA.
AC Q9KWG4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE DIPHOSPHOMEVALONATE DECARBOXYLASE.
OS Streptomyces sp. (strain CL190).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=93372;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CL190;
RX MEDLINE=20353446; PubMed=10894721;
RA Takagi M., Kuzuyama T., Takahashi S., Seto H.;
RT "A gene cluster for the mevalonate pathway from Streptomyces sp.
RT strain CL190";
RL J. Bacteriol. 182:4153-4157(2000).
DR EMBL: AB037666; BAB07791.1; -.
DR InterPro: IPR001459; -.
DR InterPro: IPR001745; -.
DR Pfam: PF00288; GHMP_Kinases; 1.
DR PRINTS: PR00959; MEVGALKINASE.
SQ SEQUENCE 350 AA; 36409 MW; 1CE61F306ACDC7CF CRC64;

Query Match 2.6%; Score 9; DB 2; Length 350;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 ALAVAAAA 19
|||||
Db 130 ALAVAAAA 138

RESULT 12
Q9YCK0 ID Q9YCK0 PRELIMINARY; PRT; 439 AA.

```

AC Q9YCK0;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE HYPOTHETICAL 45.2 KDA PROTEIN APE1257.
GN APE1257.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
DR EMBL: AP000061; BAA80247.1;
KW Hypothetical protein.
SQ SEQUENCE 439 AA; 45223 MW; 935D70CED06CF1EA CRC64;

Query Match 2.6%; Score 9; DB 1; Length 439;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 SALAVAAAA 18
| | | | | | | |
DB 398 SALAVAAAA 406

RESULT 13
Q9FRFC3
ID Q9FRFC3 PRELIMINARY; PRT; 518 AA.
AC Q9FRFC3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE PUTATIVE CYTOCHROME P450-RELATED PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,
RA Zismann V., Pal G., Bowman C.L., Fujii C.Y., VanAken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 3 BAC OSJNBa0013M12 genomic sequence."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC082644; AAC46147.1;
SQ SEQUENCE 518 AA; 58749 MW; E3B82EDD23183C81 CRC64;

Query Match 2.6%; Score 9; DB 10; Length 518;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LAVAAAAA 20
| | | | | | | |
DB 7 LAVAAAAA 15

RESULT 14

```

```

O86516
ID O86516 PRELIMINARY; PRT; 776 AA.
AC O86516;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE HYPOTHETICAL 45.2 KDA PROTEIN APE1257.
GN APE1257.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
DR EMBL: AP000061; BAA80247.1;
KW Hypothetical protein.
SQ SEQUENCE 439 AA; 45223 MW; 935D70CED06CF1EA CRC64;

Query Match 2.6%; Score 9; DB 2; Length 776;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 AAAAAAALT 23
| | | | | | | |
DB 18 AAAAAAALT 26

RESULT 15
Q9NB03
ID Q9NB03 PRELIMINARY; PRT; 840 AA.
AC Q9NB03;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE NUCLEAR RECEPTOR FTZ-F1 PROTEIN.
OS Aedes aegypti (Yellowfever mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Aedes.
OX NCBI_TaxID=7159;
RN [1]
RP SEQUENCE FROM N.A.
RA Li C., Kapitskaya M.Z., Zhu J., Miura K., Segraves W., Raikhel A.S.;
RT "Conserved Molecular Mechanism for the Stage Specificity of the
RT Mosquito Vitellogenic Response to Ecdysone."
RL Dev. Biol. 0:0-0(2000).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.
EMBL: AF274870; AAF82307.1;
DR InterPro: IPR000536;

```

DR InterPro: IPR001628; -.
 DR InterPro: IPR001723; -.
 DR Pfam: PF0104; hormone_rec; 1.
 DR Pfam: PF0105; zf-C4; 1.
 DR PRINTS; PR00398; STRDHORMONER.
 DR PRINTS; PR00047; STROIDFINGER.
 DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
 DR SMART; SM00430; HOLI; 1.
 KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
 KW Zinc-finger.
 SQ SEQUENCE 840 AA; 89623 MW; AC85DB8DCE2DDA3A CRC64;

Query Match 2.6%; Score 9; DB 5; Length 840;
 Best Local Similarity 100.0%; Pred. No. 5.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 AVAAAAAAL 21
 |||||
 Db 191 AVAAAAAAL 199

Search completed: August 23, 2001, 11:19:27
 Job time: 199 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 11:19:17 ; Search time 2905.14 Seconds
(without alignments)
3961.257 Million cell updates/sec

Title: US-09-654-652A-4
Perfect score: 744
Sequence: 1 atggttagcgcaaaaggattt.....gagatgacgaacctgtctcg 744

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733974588 residues

Total number of hits satisfying chosen parameters: 2688314

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : GenEmbl.*
1: gb_bai.*
2: gb_ba2.*
3: gb_ba3.*
4: gb_in1.*
5: gb_in2.*
6: gb_in3.*
7: gb_om.*
8: gb_ov.*
9: gb_pat1.*
10: gb_pat2.*
11: gb_ph.*
12: gb_pl1.*
13: gb_pl2.*
14: gb_pl3.*
15: gb_pl4.*
16: em_ba1.*
17: em_ba2.*
18: em_fun.*
19: em_htgo_hum.*
20: em_htgo_inv.*
21: em_htgo_rod.*
22: em_htg_hum1.*
23: em_htg_hum2.*
24: em_htg_hum3.*
25: em_htg_hum4.*
26: em_htg_hum5.*
27: em_htg_hum6.*
28: em_htg_hum7.*
29: em_htg_hum8.*
30: em_htg_inv1.*
31: em_htg_inv2.*
32: em_htg_other.*
33: em_htg_rod.*
34: em_hum1.*
35: em_hum2.*
36: em_hum3.*
37: em_hum4.*
38: em_hum5.*
39: em_hum6.*
40: em_hum7.*
41: em_in.*
42: em_om.*
43: em_or.*
```

Result No.	Query ^a			Length	DB	ID	Description
	Score	Match					
1	741	99.6	1426	3	FIBGLUC	M3676	F. succinogenes
2	50	6.7	1416	2	BSPBGAIG	Z12151	Bacillus sp.
3	49.6	6.7	1234	56	SYHA78M	Z25882	Synthetic B
4	47.4	6.4	927	9	AX000165	AX000165	Sequence
5	44.6	6.0	1228	56	SYHAOM	Z25873	Synthetic B
6	44.6	6.0	1234	56	SYHA12Y13	Z25876	Synthetic B
7	44.6	6.0	1234	56	SYHA16M	Z25878	Synthetic B
8	44.6	6.0	1234	56	SYHA2M	Z25879	Synthetic B


```

|||||
Db 937 AGAGATGACGACCTGCTCCG 957

RESULT 2
LOCUS BSPGALG 1416 bp DNA BCT 07-NOV-1994
DEFINITION Bacillus sp. bgaA gene for lichenase.
ACCESSION Z12151
VERSION 212151.1 GI:296931
KEYWORDS bga1 gene; lichenase.
SOURCE Bacillus sp.
ORGANISM Bacillus sp.
REFERENCE 1 (bases 1 to 1416)
AUTHORS Tabernero,C., Coll,P.M., Fernandez-Abalos,J.M., Perez,P. and
          Santamaria,R.I.
TITLE Cloning and DNA sequencing of bgaA, a gene encoding an
       endo-beta-1,3-1,4-glucanase, from an alkalophilic Bacillus strain
       (N137)
JOURNAL Appl. Environ. Microbiol. 60 (4), 1213-1220 (1994)
MEDLINE 94288605
REFERENCE 2 (bases 1 to 1416)
AUTHORS Santamaria,R.I.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-1992) Ramon I. Santamaria, Instituto de
       Microbiologia Bioquimica, C.S.I.C./ Universidad, de Salamanca,
       Plaza de la Merced S/N, Salamanca, Salamanca, 37008, Spain
FEATURES
       source
           1..1416
               /organism="Bacillus sp."
               /strain="137"
               /db_xref="taxon:1409"
               /clone="10"
               199..204
               -35_signal
               -10_signal
               5'UTR
               RBS
               gene
               CDS
               388..1218
               /gene="bga1"
               /gene="bga1"
               /codon_start=1
               /transl_table=11
               /product="lichenase"
               /protein_id="CAA78135.1"
               /db_xref="GI:296932"
               /db_xref="SPTREMBL:Q45648"
               /translation="MKKLVLFSCMLLLFSLGSLVSPQVAEAAERPIGTAFTVET
               ESYDSRWKAGVWSNGQMFNATWYPEQVTFSGNKMLOIEREDNETAPPYKAGELR
               TNDYHYGLFEYSMPKPAKSTGTVSFFTYTGPWDENDPDDEITFLGKDTKIQFN
               YTNQVGGNEHYHELGFDAADENTYAFERPESTPWFVNGELVHTATENIPQPKI
               MNNLPFGIVDGTGRFNGEDIPVTQYDWNKYTPLEELGCTNEKNKKYKCKTKYK
               "
               388..480
               sig_peptide
               /gene="bga1"
               481..1215
               mat_peptide
               /gene="bga1"
               /product="lichenase"
BASE COUNT 509 a 181 c 276 g 450 t
ORIGIN

Query Match 6.7%; Score 50; DB 2; Length 1416;
Best Local Similarity 51.5%; Pred. No. 0.0055;
Matches 172; Conservative 0; Mismatches 150; Indels 12; Gaps 2;

Qy 25 ggtgcgaactctacacgttagaagattcagtagcgttaagttgaagcccgtagaag 84
|||||
Db 679 GCTGGAGAACTCCGACAAACGATTTTATCATTTACGCGGTGTTTGAAGTGAGTATGAAA 738
|||||
Qy 85 atggcagccgcatcggggaacagtcagttccattctctaccagaatggttcggaatc 144

```

```

|||||
Db 739 CTTGCAAAAGTCAACGGGAACAGTCTCTTCATTTTACCTATACTGACCTTGGATTGG 798
|||||
Qy 145 gccgatggaagccctggtagaagtggatattgaagtcttcgcggaagaatccgggcagt 204
|||||
Db 799 GAGAAATGAT--CCATGGGATGAATGATCGAATTTTAGTAAAGTACTACTATAA 855
|||||
Qy 205 ttccagtcacaatcattaccggttaagccggcgagcaaaagactagcgaagacacat 264
|||||
Db 856 ATACAATTTAATTATTTTACAACAGGAGTAGGGGGAATGAGCATT-----ACCAT 906
|||||
Qy 265 qctgttagcccccgcgcgcagctttccacacctacggtctcgaatgactccgaat 324
|||||
Db 907 GAATTAGGATTTGATGACGAGATGATTTTAATACGATGCTTTTGTAGTGAGACCAGAA 966
|||||
Qy 325 taagtccgctggactgttgacggtcaggaaagtc 358
|||||
Db 967 TCTATTGCTTGGTTTGTAAATGGTGAATTGGTTC 1000
|||||

RESULT 3
SYHA78M
LOCUS SYHA78M 1234 bp DNA SYN 04-APR-2000
DEFINITION Synthetic B.macerans/B.amyloliquefaciens hybrid gene for
       endo-1,3-1,4-beta-glucanase.
ACCESSION Z25882
VERSION 225882.1 GI:398202
KEYWORDS beta-glucanase; endo-1,3-1,4-beta-glucanase.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 777)
AUTHORS Hofemeister,J., Kurtz,A., Borriss,R. and Knowles,J.
TITLE The beta-glucanase gene from Bacillus amyloliquefaciens shows
       extensive homology with that of Bacillus subtilis
       Gene 49 (2), 177-187 (1986)
JOURNAL 87192007
MEDLINE 2 (bases 396 to 852)
REFERENCE Borriss,R., Buettner,K. and Maentsaelae,P.
AUTHORS Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans:
       homologues to other beta-glucanases
       Mol. Gen. Genet. 222 (2-3), 278-283 (1990)
JOURNAL 91109712
MEDLINE 3 (bases 1 to 1234)
REFERENCE Olsen,O., Borriss,R., Simon,O. and Thomsen,K.K.
AUTHORS Hybrid Bacillus (1-3-1,4)-beta-glucanases: engineering thermostable
       enzymes by construction of hybrid genes
       Mol. Gen. Genet. 225 (2), 177-185 (1991)
JOURNAL 91172113
MEDLINE 4 (bases 1 to 1234)
REFERENCE Pollitz,O., Simon,O., Olsen,O. and Borriss,R.
AUTHORS Determinants for the enhanced thermostability of hybrid
       (1-3,1-4)-beta-glucanases
       Eur. J. Biochem. 216 (3), 829-834 (1993)
JOURNAL 94009045
MEDLINE 5 (bases 1 to 1234)
REFERENCE Pollitz,O.
AUTHORS Direct Submission
TITLE Submitted (26-AUG-1993) Pollitz O., Humboldt University of Berlin,
       Biology, Warschauer Str. 43, BERLIN, Germany, D-10243
       Location/Qualifiers
       1..1234
           /organism="synthetic construct"
           /db_xref="taxon:32630"
           /focus
           1..1234
           /organism="Bacillus amyloliquefaciens"
           /db_xref="taxon:1390"
           1..777
           /organism="Bacillus amyloliquefaciens"
           /db_xref="taxon:1390"
           1..1234
           source
FEATURES
       source
           /db_xref="taxon:32630"
           /focus
           1..1234
           /organism="Bacillus amyloliquefaciens"
           /db_xref="taxon:1390"
           1..777
           /organism="Bacillus amyloliquefaciens"
           /db_xref="taxon:1390"
           1..1234
           source

```

```

/organism="Paenibacillus macerans"
/db_xref="taxon:44252"
469. .543
469. .1188
/function="endo-1,3-1,4-beta-glucanase"
/note="Hybrid between bgla (pos. 1 to 777) and bglm
(pos.in hybrid 778 to 1234, pos. original 396 to 852)"
/codon_start=1
/transl_table=11
/label=HA78M
/evidence-experimental
/product="hybrid-endo-1,3-1,4-beta-glucanase"
/protein_id="CAA81101.1"
/db_xref="GI:398203"
/transl_table="MKRVLLILVTGLFMSLGIITSSVSAQTGGFFPNSVSLQW
KMGVSDGNMNCVTRANNVSWTSLGEMRLATSPSYNKFDGGENRSVOTYGYGLYEV
RMKPAKNTYVSSFTYGPAGHQWDEIDIEFGKDKTRVQFNITNGVGGHEKVIS
LGFDAKSGFHTYADWQPGYIKWYVDGVGLKHTATANTPSTPGKIMMNLWNGTVDDWL
GSYGANPLXYAEYDVKVTSN"
778. .1234
source
/organism="Paenibacillus macerans"
/db_xref="taxon:44252"
373 a 224 c 281 g 356 t
BASE COUNT
ORIGIN

```

```

Query Match 6.7%; Score 49.6; DB 56; Length 1234;
Best Local Similarity 51.8%; Pred. No. 0.0071;
Matches 174; Conservative 0; Mismatches 144; Indels 18; Gaps 2;

QY 13 aaggatttagcgggtgcgaactctacagcttagaagaagtcagtgaggttaagttttaa 72
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 712 AACAAAGTTTGAAGTGGGGGAAACCGCTCGGTTCACAAATATGGCTATGGACTTTATGAA 771

QY 73 gccctatgaagataggcagccgcagtcaggaacagtcagttccatgttctctaccagaat 132
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 772 GTCAGAAATGAAGCCAGCCAAAATAACAGAAATGTCTCATCTCTTTTTCACGTA----- 824

QY 133 ggttcgaaatccgcagatgaagccctgggttagaagtgatgaggtatgaaagtcaggaag 192
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 825 --TACAGGACCTGCTCATGCGCACAAATGGGATGAATAGATATGAAATTTCTAGGAAA 882

QY 193 aatccgggcagtttcagtcacacatcattaccgtaagcgccgcgcacaaagactagc 252
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 883 GACAGCAGAAAAGTCCAGTTAACTATTATACCAATGGGTTGGCGGTCTAT----- 933

QY 253 gaaagcaccatgctgttagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 312
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 934 GAAAAGGTTATCTCTTGGCTTTGATGATCAAAAGGCTTCCATACCTATGCTTTCGAT 993

QY 313 tggactccgaattacgtccgctggactgtgacggt 348
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 994 TGGCAGCAGGATATTAATGGTATGTAGACGGT 1029

```

```

RESULT 4
AX000165 LOCUS AX000165 927 bp DNA PAT 10-MAR-2000
DEFINITION Sequence 2 from Patent WO9906573.
ACCESSION AX000165
VERSION AX000165.1 GI:7240678
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 927)
AUTHORS Maurer,K. and Hillen,W.
TITLE NEW BETA-GLUCANASE FROM A BACILLUS
JOURNAL Patent: WO 9906573-A 2 11-FEB-1999;
MAURER KARL HEINZ (DE); HENKEL KGA (DE)
FEATURES
location/Qualifiers
source 1. .927
/organism="unidentified"

```

```

/strain="BACILLUS ALKALOPHILUS DSM 9956"
/db_xref="taxon:32644"
343 a 150 c 218 g 216 t
BASE COUNT
ORIGIN

Query Match 6.4%; Score 47.4; DB 9; Length 927;
Best Local Similarity 50.7%; Pred. No. 0.028;
Matches 173; Conservative 0; Mismatches 156; Indels 12; Gaps 2;

QY 16 gatttagcgggtccgaaactctacagcttagaagaagtcagtcagtgtaagtttgaagcc 75
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 286 GAATATAAGGCTGGGGAATTAAGAAAGCAATCAATGTTTATCAATACGGGTGTTTGAAGTC 345

QY 76 gatatgaagatggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 135
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 346 ATATGAAGCCAGCGAAATCAACAGAACCGTCTCTTCACTCTTTACATATACGGGTCCA 405

QY 136 tccgaaatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 195
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 406 TGGGATTGGGATAATG---ATCCTTTGGGATGAATCGATATTGAGTCTCTTGGAAAGGAT 462

QY 196 cgggcagtttccagtcacacatcattaccggtgaagccgcgcgcgcgcgcgcgcgcgcgc 255
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 463 ACACAAGAGTCCCAATTTAACTATTTTACTAA-----CGGAGTAGGAACAATGAA 513

QY 256 aagcaccatgctgttagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 315
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 CATTACCAAGAAATAGGTTTCGATCGATCAGAACTCTTTTAATACGTATGCTTTTGAATGG 573

QY 316 actcgaattacgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 356
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 574 AGACCAGAATCAATAGTTAGTGTACGTAAACGGAGAAATAGT 614

RESULT 5
SYHAOM LOCUS SYHAOM 1228 bp DNA SYN 04-APR-2000
DEFINITION Synthetic B.macerans/B.amyloliquefaciens hybrid gene for
endo-1,3-1,4-beta-glucanase.
ACCESSION Z25873.1 GI:398185
VERSION Z25873.1
KEYWORDS beta-glucanase; endo-1,3-1,4-beta-glucanase.
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 543)
AUTHORS Hofemeister,J., Kurtz,A., Borriess,R. and Knowles,J.
TITLE The beta-glucanase gene from Bacillus amyloliquefaciens shows
extensive homology with that of Bacillus subtilis
JOURNAL Gene 49 (2), 177-187 (1986)
MEDLINE 87192007
REFERENCE 2 (bases 168 to 852)
AUTHORS Borriess,R., Buettner,K. and Maentsaelae,P.
TITLE Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans:
homologies to other beta-glucanases
JOURNAL Mol. Gen. Genet. 222 (2-3), 278-283 (1990)
MEDLINE 91109712
REFERENCE 3 (bases 1 to 1228)
AUTHORS Politz,O., Simon,O., Olsen,O. and Borriess,R.
TITLE Determinants for the enhanced thermostability of hybrid
(1-3-1-4)-beta-glucanases
JOURNAL Eur. J. Biochem. 216 (3), 829-834 (1993)
MEDLINE 94009045
REFERENCE 4 (bases 1 to 1228)
AUTHORS Politz,O., Simon,O., Olsen,O. and Borriess,R.
TITLE Determinants for the enhanced thermostability of hybrid
(1-3-1-4)-beta-glucanases
JOURNAL Eur. J. Biochem. (1993) In press
REMARK (sites)
AUTHORS 5 (bases 1 to 1228)
POLITZ,O.
TITLE Direct Submission

```

JOURNAL Submitted (26-AUG-1993) Politz O., Humboldt University of Berlin,
Biology, Warschauer Str. 43, BERLIN, Germany, D-10243

FEATURES
source 1..1228
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /focus
source 1..1228
 /organism="Bacillus amyloliquefaciens"
 /db_xref="taxon:1390"
source 1..543
 /organism="Bacillus amyloliquefaciens"
 /db_xref="taxon:1390"
source 1..1228
 /organism="Paenibacillus macerans"
 /db_xref="taxon:44252"
 /organism="Paenibacillus macerans"
 /db_xref="taxon:44252"
sig_peptide 469..543
CDS 469..1182
 /function="endo-1,3-1,4-beta-glucanase"
 /note="Hybrid between bgla (Pos. 1 to 543) and bglm
 (Pos.in hybrid 544 to 1228, Pos. Original 168 to 852),"
 /codon_start=1
 /transl_table=11
 /label=HAOM
 /evidence=experimental
 /product="hybrid-endo-1,3-1,4-beta-glucanase"
 /protein_id="CAA81092.1"
 /db_xref="GI:398186"
translation="MKRVLLILVLTGLFMSLCGTTSSVSAGSFWEPLSPENPSTWPKA
DYSGNGVNCNTRANNVNFTNDKLUKLGTLSAYNKFCAEYRNTIYGIGLYEIVSM
FKAPKTIYSVFYTGPAHQWDIEFLGKDTTFQFNYYTNGVGSHGVSLG
FPASKGHTFYADWPQGYIKWYVDGVLKTKATANIPSTFGKIMMNLWGVTGDWLIGS
YGANPLYAEYDWVKYTSN"
544..1228
 /organism="Paenibacillus macerans"
 /db_xref="taxon:44252"
BASE COUNT 371 a 220 c 290 g 347 t
ORIGIN

Query Match 6.0%; Score 44.6; DB 56; Length 1228;
Best Local Similarity 52.9%; Pred. No. 0.17;
Matches 154; Conservative 0; Mismatches 119; Indels 18; Gaps 2;

Qy 58 tacggttaagttgaagcccgatgaagatggcagccgcatacgggaacagtcagtccatg 117
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 751 TACGGCGCTGTACGAGGTCACTATGAACGCACGCCAAAATAACAGGAATTGCTCATCCTTT 810
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 118 ttctctaccagaatggttcccgaatgccgatggaggccctgggttagaagtgatat 177
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 811 TTCACGTA-----TACAGGACCTGCTCATGGCACAAATGGGATGAATAGATATC 861
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 178 gaagttctcggaagaatccggcgagtttccagtcacaacatattaccggttaagccgcg 237
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 862 GAATTTC TAGGAAAGACACGCAAAAGTCCAGCTTTAACTATTATACCAATGGGGTTGGC 921
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 238 gcacaaaagactagcaaaagaccacatgctgtttagcccccgccgcatcaggtcttccac 297
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 922 GTTCAT-----GAAAAGTTATCTCTCTGGCTTTTATGCATCAAGGGCTTCAT 972
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 298 acctacggtctcgaatggactccgaattaacgtccgctggagctgttacggt 348
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 973 ACCTATGCTTTTCGATTGGCAGCCAGGGGTATATTAAATGGTATGTAGACGCT 1023
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
SYHA12Y13
LOCUS SYHA12Y13 1234 bp DNA SYN 04-APR-2000
DEFINITION Synthetic B.macerans/B.amyloliquefaciens hybrid gene for
 endo-1,3-1,4-beta-glucanase.
ACCESSION Z25876
VERSION Z25876.1 GI:398190
KEYWORDS beta-glucanase; endo-1,3-1,4-beta-glucanase.

SOURCE	ORGANISM	synthetic construct.
REFERENCE	AUTHORS	synthetic construct
TITLE		artificial sequence.
JOURNAL		1 (bases 1 to 579)
MEDLINE		Hofemeister,J., Kurtz,A., Borriass,R. and Knowles,J.
REFERENCE	AUTHORS	The beta-glucanase gene from <i>Bacillus amyloliquefaciens</i> shows
TITLE		extensive homology with that of <i>Bacillus subtilis</i>
JOURNAL		Gene 49 (2), 177-187 (1986)
MEDLINE		87192007
REFERENCE	AUTHORS	2 (bases 198 to 852)
TITLE		Borriass,R., Buettnier,K. and Maentsaelae,P.
JOURNAL		Structure of the beta-1,3-1,4-glucanase gene of <i>Bacillus macerans</i> :
MEDLINE		homologies to other beta-glucanases
REFERENCE	AUTHORS	Mol. Gen. Genet. 222 (2-3), 278-283 (1990)
TITLE		91109712
JOURNAL		3 (bases 1 to 1234)
MEDLINE		Pollitz,O., Simon,O., Olsen,O. and Borriass,R.
REFERENCE	AUTHORS	Determinants for the enhanced thermostability of hybrid
TITLE		(1-3,1-4)-beta-glucanases
JOURNAL		Eur. J. Biochem. 216 (3), 829-834 (1993)
MEDLINE		94009045
REFERENCE	AUTHORS	4 (bases 1 to 1234)
TITLE		Pollitz,O., Simon,O., Olsen,O. and Borriass,R.
JOURNAL		Determinants for the enhanced thermostability of hybrid
MEDLINE		(1-3,1-4)-beta-glucanases
REFERENCE	AUTHORS	Eur. J. Biochem. (1993) In press
TITLE		(sites)
JOURNAL		5 (bases 1 to 1234)
MEDLINE		Pollitz,O.
REFERENCE	AUTHORS	Direct Submission
TITLE		Submitted (26-AUG-1993) Pollitz O., Humboldt University of Berlin,
JOURNAL		Biology, Warschau Str. 43, BERLIN, Germany, D-10243
MEDLINE		Location/Qualifiers
FEATURES	source	1. .1234
		/organism="synthetic construct"
		/db_xref="taxon:32630"
	source	/focus
		1. .1234
		/organism="Bacillus amyloliquefaciens"
		/db_xref="taxon:1390"
	source	1. .579
		/organism="Bacillus amyloliquefaciens"
		/db_xref="taxon:1390"
	source	1. .1234
		/organism="Paenibacillus macerans"
		/db_xref="taxon:44252"
	sig_peptide	469. .543
	CDS	469. .1188
		/function="endo-1,3-1,4-beta-glucanase"
		/notes="Hybrid between bgla (Pos. 1 to 579) and bglm
		(Pos. Hybrid 580 to 1234, Pos. Original 198 to 852)
		Deletion of Y13 of HA12-M"
		/codon_start=1
		/transl_table=11
		/label=HA12M
		/evidence=experimental
		/product="hybrid-endo-1,3-1,4-beta-glucanase"
		/protein_id="CA81095.1"
		/db_xref="GI:398191"
		/translation="MKRVLLILVTGLFMSLCGITSVSQAQTGSGFEPFNSRNPSTW
		KADYGGNGVFNCTWRANNVFNFDKLGKLTSSAYNKFDCAEYRSTNIYGYLVE
		SMKPAKNTIVTSFFTYTQTPAGHGTQWDEIDIEFLGKDTTKVQFNYYTNGVGGRKVIS
		LGFADASKGFHTYAFQWQPGYIKWYVDGLKHTATANIPTPGKIMNLANWGTCVDDWL
		GSYNGANPLIAEYDWMKYTSN"
		580. .1234
	source	/organism="Paenibacillus macerans"
		/db_xref="taxon:44252"
	BASE COUNT	374 a 225 c 287 g 348 t
	ORIGIN	

Query Match 6.0%; Score 44.6; DB 56; Length 1234;
Best Local Similarity 52.9%; Pred. No. 0.17;


```

/label=HA36M
/evidence=experimental
/product="hybrid-endo-1,3-1,4-beta-glucanase"
/protein_id="CAA81099.1"
/db_xref="db:398199"
/translation="MKRVLLILVTGLFMSLCGITSVSQAQTGSGSFPEPNSYNSGLWQ
KADGYSDDGMFNCFTWRANVNFNDGKLGKLTSSAYNKPDCAEYRSTNIYGYGLYEV
SKRPAGKNTGISVSFFTYTPAIGTHQWDEIFELGKDTTKVQFNYYTNGVGGHEKIVS
LGFDAKSGFYTAFAWQPGYIKWYVDGLVKHTATANIPSTPGKIMNMLNNGTGVDDDL
GSYNGANPLYAEVDWVKYTSN"
652. 1234
/organism="Paenibacillus macerans"
/db_xref="taxon:44252"
376 a 225 c 281 g 352 t
BASE COUNT
ORIGIN

```

Query Match	6.08;	Score 44.6;	DB 56;	Length 1234;
Best Local Similarity	52.9%;	Pred. No. 0.17;		
Matches 154;	Conservative	0;	Mismatches 119;	Indels 18; Gaps 2;
QY	58	tacggtaagtttaagcccgctatgaagatgacgcgcgcacgtcgggaacagtcagttcccatg	117	
Db				
Db	757	TACGGCCTGTACGAGGTCAGTATGAAGCCACGCCAAAATACAGGAATGTCTCATCCTTT	816	
QY	118	ttctctacagaatgggttcggaatccgcgatggaagggccctgggtagaagtgatgataatt	177	
Db				
Db	817	TTCAACGTA-----TACAGGACCTGCTCATGGCACACAATCGGATGAATAGATATC	867	
QY	178	gaagttctcggcaagaatccgggcagtttccagtcccaacatcattaccggtlaagggccggc	237	
Db				
Db	868	GAATTTCTAGAAAAGACACGACAAAAGTCACGTTTAACATTATTACCAATGGGTTGGC	927	
QY	238	gcacaaaagactagcaaaagaccatgctgttagcccccgcgcgcgatcagagctttccac	297	
Db				
Db	928	GGTCAT-----GAAAGGTTATCTCTCTGGCTTTGATGTCATCAAGAGGGTTCAT	978	
QY	298	acctacggtctcgaatggactccgaattaccgtccgctggactgttgacggt	348	
Db				
Db	979	ACCTATGCTTTCCGATTCGGACGCCAGGCTATATTAAATGGTATGTAGACGGT	1029	

RESULT	10
LOCUS	SYHA4M 1234 bp DNA SYN 04-APR-2000
DEFINITION	Synthetic B.macerans/B.amyloliquefaciens hybrid gene for endo-1,3-1,4-beta-glucanase.
ACCESSION	Z25881.1 GI:399200
VERSION	Z25881
KEYWORDS	beta-glucanase; endo-1,3-1,4-beta-glucanase.
SOURCE	synthetic construct
ORGANISM	artificial sequence.
REFERENCE	1 (bases 1 to 555)
AUTHORS	Hofemeister,J., Kurtz,A., Borriss,R. and Knowles,J.
TITLE	The beta-glucanase gene from Bacillus amyloliquefaciens shows extensive homology with that of Bacillus subtilis Gene 49 (2), 177-187 (1986)
JOURNAL	87192007
MEDLINE	2 (bases 174 to 852)
REFERENCE	Borriss R., Buettner,K. and Maentzsaetae,P.
AUTHORS	Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans homologues to other beta-glucanases Mol. Gen. Genet. 222 (2-3), 278-283 (1990)
JOURNAL	91109712
MEDLINE	3 (bases 1 to 1234)
REFERENCE	Poltitz,O., Simon,O., Olsen,O. and Borriss,R.
AUTHORS	Determinants for the enhanced thermostability of hybrid (1-3,1'-4)-beta-glucanases Eur. J. Biochem. 216 (3), 829-834 (1993)
JOURNAL	94009045
MEDLINE	4 (bases 1 to 1234)
REFERENCE	Poltitz,O., Simon,O., Olsen,O. and Borriss,R.
AUTHORS	

TITLE	Determinants for the enhanced thermostability of hybrid (1-3,1-4)-beta-glucanases
JOURNAL	Eur. J. Biochem. (1993) In press
REMARK	(sites)
REFERENCE	5 (bases 1 to 1234)
AUTHORS	Politz,O.
TITLE	Direct Submission
JOURNAL	Submitted (26-AUG-1993) Politz O., Humboldt University of Berlin, Biology, Warschauer Str. 43, BERLIN, Germany, D-10243
FEATURES	Location/Qualifiers
source	1. .1234 /organism="synthetic construct" /db_xref="taxon:32630"
source	1. .1234 /focus /organism="Bacillus amyloliquefaciens" /db_xref="taxon:1390"
source	1. .555 /organism="Bacillus amyloliquefaciens" /db_xref="taxon:1390"
source	1. .1234 /organism="Paenibacillus macerans" /db_xref="taxon:44252"
sig_peptide	469. .543
CDS	469. .1188 /function="endo-1,3-1,4-beta-glucanase" /note="Hybrid between bgla (Pos. 1 to 555) and bglm (Pos.in hybrid 556 to 1234, Pos. original 174 to 852)" /codon_start=1 /transl_table=11 /label=HA4M /evidence=experimental /product="hybrid-endo-1,3-1,4-beta-glucanase" /protein_id="CA811100.1" /db_xref="GI:398201"
source	/translation="MKRVLLIVLTGFMSCGITSVSQAOTGGVFWPEPLSYFNPSTWEE SKNGVGGVFNCTWRANVNFTNDGKLGLTSAYNKFDCAEYRSTNIYVGLYEV SNKPAKNTYISVSFFTYTQPHAGTOWDEIDIEFLGKDTTKVQNYVTNGVGGHEKVIS LFGDASKGFLYAFDMQPGYIKWYVDGLVKHTATANIPTSPGKIMNLNWNGTGVDDDL GSVNGANPLAYEDWVKYTSN" 556. .1234 /organism="Paenibacillus macerans" /db_xref="taxon:44252"
BASE COUNT	375 a 223 c 290 g 346 t
ORIGIN	

Query Match	6.0%;	Score	44.6;	DB	56;	Length	1234;
Best Local Similarity	52.9%;	Pred.	No.	0.17;			
Matches	154;	Conservative	0;	Mismatches	119;	Indels	18;
Gaps	2;						
Qy	58	tacggtaaagttaaagcccgatgaagatgcagcgcctcggagacagtcagttccatg	117				
Db	757	TAGGCGCTGTACGAGGTCAATGTGAAGCAGCCAAAATACAGGAATTGTCATCCTTT	816				
Qy	118	ttcctctaccagaatggttcggaatcgcgcatggaagccctgggtagaagtggatatt	177				
Db	817	TTCCAGTA-----TACAGGACCTGCTCATGGCACACAATGGGATGAATAGATATC	867				
Qy	178	gaagtctcggcgaagaatcgggcagtttccagttccaatccaaatcattaccgtaagccgcg	237				
Db	868	GAATTTCTAGGAAACACACGACAAAGTCAGTTTAACTATTATACCAATGGGGTTGGC	927				
Qy	238	gcacaaagactacgaaaagaccattgctgttagccgcgcgcgcgactcaggtttccac	297				
Db	928	GGTCAT-----GAAAAGGTTATCTCTCTTGGCTTTGCATGCATCAAAAGGGCTTCCAT	978				
Qy	298	acctacggtctcgaatggactccgaattacgctcgcgtggactgtgaacggt	348				
Db	979	ACGTATGCTTTTCGATTGGCAGCCAGGGTATTAAAAATGTTATGTAGACGGT	1029				

QY	58	tacgttaagtgaagcccgatgaagatggcagccgcatcggaacagtcagttcccatg	117
Db	757	TACGGCGCTGTACGAGGTGACGTAGATGAAGCAGCAGCAAAAATACAGGAATGTCTCATCCTTT	816
QY	118	ttcctctaccagaatggtttccgaatacgcgcgatggaagccctgggtagaagtggatatt	177
Db	817	TTCACGTA-----TACAGGACCTGCTCATGGCACAAATGGGATCAATAGATATC	867
QY	178	gaagttctcggcaagaatccgggcagtttccagttcccaacatcattaccggtaagccgcgc	237
Db	868	GAATTTCTTAGGAAAGACACGACAAAGTCCAGTTTAACATATTATACCAATGGGCTTGGC	927
QY	238	gcacaaagacagtagcgaagaacaccatgcgtgtagccgcgcgcgatcaggctttccac	297
Db	928	GGTCAT-----GAAAGGTTATCTCTCTGGCTTTGATGCATCAAGGGCTTCCAT	978
QY	298	acctcaggtctcgaatggactccgaattacgtccgctggagctgttgacggt	348
Db	979	ACCTATGCTTTTCGATTGGCAGCCAGGCTATATTAAATGGTATGTAGACGGT	1029

Best Local Similarity 52.9%; Pred. NO. 0.17;
Matches 154; Conservative 0; Mismatches 119; Indels 18; Gaps

SYHA12MX	1237 bp	DNA	SYN	04-APR-2000
LOCUS	Synthetic B.macerans/B.amyloliquefaciens hybrid gene for			
DEFINITION	endo-1,3-1,4-beta-glucanase.			
ACCESSION	Z25875			
VERSION	Z25875.1 GI:406780			
KEYWORDS	endo-1,3-1,4-beta-glucanase.			
SOURCE	synthetic construct.			
ORGANISM	artificial sequence.			
REFERENCE	1 (bases 1 to 579)			
AUTHORS	Hofemeister, J., Kurtz, A., Borriss, R. and Knowles, J.			
TITLE	The beta-glucanase gene from Bacillus amyloliquefaciens shows extensive homology with that of Bacillus subtilis			
JOURNAL	Gene 49 (2), 177-187 (1986)			
MEDLINE	R7192007			
REFERENCE	2 (bases 196 to 852)			
AUTHORS	Borriss, R., Buettner, K. and Maentsaelae, P.			
TITLE	Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans homologues to other beta-glucanases			
JOURNAL	Mol. Gen. Genet. 222 (2-3), 278-283 (1990)			
MEDLINE	91109712			
REFERENCE	3 (bases 1 to 1237)			
AUTHORS	Pollitz, O., Simon, O., Olsen, O. and Borriss, R.			
TITLE	Determinants for the enhanced thermostability of hybrid (1-3,1-4)-beta-glucanases			
JOURNAL	Eur. J. Biochem. (1993) In press			
REMARK	(sites)			
REFERENCE	4 (bases 1 to 1237)			
AUTHORS	Pollitz, O.			
TITLE	Direct Submission			
JOURNAL	Submitted (26-AUG-1993) Pollitz, O., Humboldt University of Berlin, Biology, Warschauer Str. 43, BERLIN, Germany, D-10243			
COMMENT	On Oct 8, 1993 this sequence version replaced gi:398189.			
FEATURES	Location/Qualifiers			
source	1..1237			
	/organism="synthetic construct"			
	/db_xref="taxon:32630"			
	/focus			
source	1..1237			
	/organism="Bacillus amyloliquefaciens"			
	/db_xref="taxon:1390"			
source	1..579			
	/organism="Bacillus amyloliquefaciens"			
	/db_xref="taxon:1390"			
source	1..1237			
	/organism="Paenibacillus macerans"			
	/db_xref="taxon:44252"			

```
sig_peptide 469. 543
CDS         469. 1191
function="endo-1,3-1,4-beta-glucanase"
/feature="hybrid between bgla (Pos. 1 to 579) and bglm
/pos.in hybrid 580 to 1237, Pos. original 196 to 852)"
/codon_start=1
/transl_table=11
/label="HA12M"
/evidence="experimental"
/product="hybrid-endo-1,3-1,4-beta-glucanase"
/protein_id="CA81094.1"
/db_xref="GI:406781"
/translation="MKRVLLIIVTGLFMSLCITSSVSAQTGGSEFFENSPSTW
EKADGYSNGVENCFTNRANNVNTDNLKGLTSSAYKPKDCAEYRSTNIYGLYE
VSMKPAKKTGIVSSFTVTGPAHQDQWDEIDIEFLGKDTKVOFNXYTNGVGGHEKI
SLGPDASGFHTYADWQPGYIKWYVDGLKHTATANIPSTPGKIMNLMWNGTGVDWW
LGSYNGANPLAEDYDWVYITSN"
source      580. 1237
/organism="Paenibacillus macerans"
/db_xref="taxon:44252"
BASE COUNT 375 a 225 c 287 g 350 t
ORIGIN

Query Match      6.0%; Score 44.6; DB 56; Length 1237;
Best Local Similarity 52.9%; Pred. No. 0.17;
Matches 154; Conservative 0; Mismatches 119; Indels 18; Gaps 2;

QY 58 tacggtaagtgaagcccgatgaagatggcagccgcatcggaacagtcagttccatg 117
||||| || || || || || || || || || || || || || || || || || || ||
Db 760 TAGCGGCTGTACGAGTCAATGATAGCGCAGCAAAAATACAGGAATTGCTCATCCTTT 819
||||| || || || || || || || || || || || || || || || || || || ||
QY 118 ttctctaccagaatgttccaaaatcccgatggaagccctggtagaagtgaatt 177
||||| || || || || || || || || || || || || || || || || || || ||
Db 820 TTCACGTA-----TACAGACCTGCTCATGCGCACACAATGGATGAAATAGATATC 870
||||| || || || || || || || || || || || || || || || || || || ||
QY 178 gaagttctcggaagaatccgggcagtttccagtcacacatcattaccggttaagccgcg 237
||||| || || || || || || || || || || || || || || || || || || ||
Db 871 GAATTTCTAGGAAACACAGCAGCAAAAGTCCAGTTTAACATATTATACCAATGGGTTGCG 930
||||| || || || || || || || || || || || || || || || || || || ||
QY 238 gacaaaagactagcgaagaaacacatgctgttagcccgccgacagctttccac 297
||||| || || || || || || || || || || || || || || || || || || ||
Db 931 GGTCAAT-----GAAAGGTTATCTCTCTGCTTTGCTTTCATCAAGGCTTCCAT 981
||||| || || || || || || || || || || || || || || || || || || ||
QY 298 acctcaggtctgaatggactcgaattacgtccgctgagctgttgacggt 348
||||| || || || || || || || || || || || || || || || || || || ||
Db 982 ACCTATGCTTTCGATTGGCAGCAGCGGTATATTAAATGTTATGTAGACCGT 1032
||||| || || || || || || || || || || || || || || || || || || ||

RESULT 13
LOCUS      BACBGLBB      1541 bp      DNA      BCT      26-APR-1993
DEFINITION B. brevis endo-beta-1,3-1,4-glucanase (bglBB) gene, complete cds.
ACCESSION  M84339
VERSION     M84339.1 GI:142583
KEYWORDS   endo-beta-1,3-1,4-glucanase.
SOURCE     Bacillus brevis (strain Alk36) DNA.
ORGANISM   Brevibacillus brevis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Brevibacillus.
            1 (bases 1 to 1541)
REFERENCE   Reid,S.J., Louw,M.E. and Watson,T.
AUTHORS    Reid,S.J., Louw,M.E. and Watson,T.
TITLE      Cloning, characterization and sequencing of a thermostable
            endo-1-3, 1-4 beta-glucanase gene from Bacillus brevis
JOURNAL    Unpublished (1991)
FEATURES   Location/Qualifiers
            source
            1. .1541
            /organism="Brevibacillus brevis"
            /strain="Alk36"
            /db_xref="taxon:1393"
            407. .1186
            /gene="bglBB"
            407. .1186
            CDS

sig_peptide 469. 543
CDS         469. 1191
function="endo-1,3-1,4-beta-glucanase"
/feature="hybrid between bgla (Pos. 1 to 579) and bglm
/pos.in hybrid 580 to 1237, Pos. original 196 to 852)"
/codon_start=1
/transl_table=11
/label="HA12M"
/evidence="experimental"
/product="hybrid-endo-1,3-1,4-beta-glucanase"
/protein_id="CA81094.1"
/db_xref="GI:406781"
/translation="MKRVLLIIVTGLFMSLCITSSVSAQTGGSEFFENSPSTW
EKADGYSNGVENCFTNRANNVNTDNLKGLTSSAYKPKDCAEYRSTNIYGLYE
VSMKPAKKTGIVSSFTVTGPAHQDQWDEIDIEFLGKDTKVOFNXYTNGVGGHEKI
SLGPDASGFHTYADWQPGYIKWYVDGLKHTATANIPSTPGKIMNLMWNGTGVDWW
LGSYNGANPLAEDYDWVYITSN"
source      580. 1237
/organism="Paenibacillus macerans"
/db_xref="taxon:44252"
BASE COUNT 375 a 225 c 287 g 350 t
ORIGIN

Query Match      5.8%; Score 43.2; DB 2; Length 1541;
Best Local Similarity 49.2%; Pred. No. 0.4;
Matches 179; Conservative 0; Mismatches 173; Indels 12; Gaps 2;

QY 7 agcgcaaggtatttagcgggtgccaaactctacacgttagaagaagttcagtcaggaag 66
||||| || || || || || || || || || || || || || || || || || || ||
Db 674 AGTGCTAGAAACTATAAGCAGGAGAGCTTCGTACCAATGATTTCTATCATTTGGACTC 733
||||| || || || || || || || || || || || || || || || || || || ||
QY 67 tttagaagcccgctgaagaatggcagccgcatcggaacagtcagttccatgttccctctac 126
||||| || || || || || || || || || || || || || || || || || || ||
Db 734 TTGTGAAGTCAGTATGAAGCTCGGAAGGTAGAAGGACCGCTGCATCCTCTTTTACCTAC 793
||||| || || || || || || || || || || || || || || || || || || ||
QY 127 cagaatggttcggaataatccgatggaagccctggtagaagtgatattgaaagtctc 186
||||| || || || || || || || || || || || || || || || || || || ||
Db 794 ACAGGGGAATGGGA---TTGGGATGGAGATCCTTGGGATGAAATTCATATTGAGTTCTTA 850
||||| || || || || || || || || || || || || || || || || || || ||
QY 187 ggaagaatccgggcagtttccagtcacacatcattaccggttaagccgacgcaaaaag 246
||||| || || || || || || || || || || || || || || || || || || ||
Db 851 GGAAGGACACGACGAGAAATACAAATTTAATTACTTTACAAATGGAGTAGGAGGAAT--- 907
||||| || || || || || || || || || || || || || || || || || || ||
QY 247 actagcgaagaacacacatgctgttagcccgccgcgatacaggctttccacacactacgt 306
||||| || || || || || || || || || || || || || || || || || || ||
Db 908 -----GAATTTTACTATGATTTAGGGTTTGTATGATCATGAGTCATTTTAATACGTATGCC 961
||||| || || || || || || || || || || || || || || || || || || ||
QY 307 ctgcgaatggactccgaattacgtccgctggaactgttgacggtcaggaagtcgcgcaagacg 366
||||| || || || || || || || || || || || || || || || || || || ||
Db 962 TTTGAATGGAGAGAGGATTCCATTACCTGTTATTAATGGAGAACGGTTTCATACAGCG 1021
||||| || || || || || || || || || || || || || || || || || || ||
QY 367 gaag 370
Db 1022 ACAG 1025

RESULT 14
LOCUS      BMBGLUC      852 bp      DNA      BCT      07-SEP-1993
DEFINITION B. macerans gene for beta-1,3-1,4-glucanase.
ACCESSION  X55959
VERSION     X55959.1 GI:296715
KEYWORDS   beta-1,3-glucanase; beta-1,4-glucanase.
SOURCE     Paenibacillus macerans.
ORGANISM   Paenibacillus macerans
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Paenibacillus.
            1 (bases 1 to 852)
REFERENCE   Bovris,R., Buettner,K. and Maentsaelae,P.
AUTHORS    Bovris,R., Buettner,K. and Maentsaelae,P.
TITLE      Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans:
            homologies to other beta-glucanases
JOURNAL    Mol. Gen. Genet. 222 (2-3), 278-283 (1990)
FEATURES   Location/Qualifiers
            source
            91109712
```



```

source      1. 852
/organism="Paenibacillus macerans"
/db_xref="taxon:44252"
93...806
CDS
/codon_start=1
/transl_table=11
/product="1,3-1,4-glucanase"
/protein_id="CAA39426.1"
/db_xref="GI:296716"
/db_xref="SWISS-PROT:P23904"
/translation="MKKSCFTLVTFAPLSFVSALAGSVFWEPLSYFNSTWEKA
DYGNGVNCFTWRANNVFTNDGKLGLTSSAYNKPDCAEYRSTNIYGYLYEVS
KPAKNTGIVSSFTYTGPAHGTQWDEIDIEFLGKDTTKVQFNYYNGVGHKVISLG
FASGPHYAFDMQPGYIKWYVDGVLKHTATANIPSTPGKIMNMLNNGTGVDDWUGS
YNGANPLYAEYDMVKYTSN"
BASE COUNT 248 a 156 c 203 g 245 t
ORIGIN

Query Match      5.8%; Score 43; DB 2; Length 852;
Best Local Similarity 52.6%; Pred. No. 0.46;
Matches 153; Conservative 0; Mismatches 120; Indels 18; Gaps 2;

QY 58 tacgtaagttaagccgqatgaagatggcagccgcatcggaacagtcagttccatg 117
||||| || || || || || || || || || || || || || || || || || || ||
Db 375 TACGGCGCTGACGAGTCAGTATGAAGCCAGCAAAATACAGGAATGTGTCATCCTTT 434
||||| || || || || || || || || || || || || || || || || || || ||

QY 118 ttctctaccagaatggttcggaatcgccgcatggaagccctgggtagaagtggatatt 177
||||| || || || || || || || || || || || || || || || || || || ||
Db 435 TTCACGTA-----TACAGACCTGTCATCGGCACACATGGATGAAATAGATATC 485
||||| || || || || || || || || || || || || || || || || || || ||

QY 178 gaagttctcggaagaatccgggcagttccagtcaccaatcattaccggtaagccgcgc 237
||||| || || || || || || || || || || || || || || || || || || ||
Db 486 GAATTCTTAGAAAGACACAGCAAAAGTCAGCTTTAACTATTATACCAATGGGTTGGC 545
||||| || || || || || || || || || || || || || || || || || || ||

QY 238 gcacaaagactagcgaagaccatcggtgttagccgcgcgcgcgcgcgcgcgcgc 297
||||| || || || || || || || || || || || || || || || || || || ||
Db 546 GGCAT-----GAAAGAGTTATCTCTCTTGGCTTTGATGCATCAAGGCGTCCCAT 596
||||| || || || || || || || || || || || || || || || || || || ||

QY 298 acctcagctcgaatggactccgaattacgcctgcgtgacttgaacggt 348
||||| || || || || || || || || || || || || || || || || || || ||
Db 597 ACCTATGCTTCGATGGCAGCCAGGGTATATTAAATGGTATGACCGT 647
||||| || || || || || || || || || || || || || || || || || || ||

RESULT 15
SYHA107M
LOCUS      SYHA107M      1234 bp      DNA      SYN      04-APR-2000
DEFINITION Synthetic B.macerans/B.amyloliquefaciens hybrid gene for
            endo-1,3-1,4-beta-glucanase.
ACCESSION 225874.1 GI:398187
VERSION   225874
KEYWORDS  beta-glucanase; endo-1,3-1,4-beta-glucanase.
SOURCE    synthetic construct.
ORGANISM  synthetic construct.
            artificial sequence.
REFERENCE 1 (bases 1 to 864)
AUTHORS   Hofemeister,J., Kurtz,A., Borriass,R. and Knowles,J.
TITLE     The beta-glucanase gene from Bacillus amyloliquefaciens shows
            extensive homology with that of Bacillus subtilis
JOURNAL   Gene 49 (2), 177-187 (1986)
MEDLINE   87192007
REFERENCE 2 (bases 1 to 1234)
AUTHORS   Borriass,R., Olsen,O., Thomsen,K.K. and von Wettstein,D.
TITLE     Hybrid bacillus endo-(1-3,1-4)-beta-glucanases: construction of
            recombinant genes and molecular properties of the gene products
JOURNAL   Carlsberg Res. Commun. 54 (2), 41-54 (1989)
MEDLINE   89374544
REFERENCE 3 (bases 463 to 852)
AUTHORS   Borriass,R., Buettner,K. and Maentsaelae,P.
TITLE     Structure of the beta-1,3-1,4-glucanase gene of Bacillus macerans:
            homologies to other beta-glucanases
JOURNAL   Mol. Gen. Genet. 222 (2-3), 278-283 (1990)
MEDLINE   91109712

```

```

REFERENCE      4 (bases 1 to 1234)
AUTHORS        Politz,O., Simon,O., Olsen,O. and Borriass,R.
TITLE          Determinants for the enhanced thermostability of hybrid
            (1-3,1-4)-beta-glucanases
JOURNAL        Eur. J. Biochem. 216 (3), 829-834 (1993)
MEDLINE        94009045
REFERENCE      5 (bases 1 to 1234)
AUTHORS        Politz,O.
TITLE          Direct Submission
JOURNAL        Submitted (26-AUG-1993) Politz O., Humboldt University of Berlin,
            Biology, Warschauer Str. 43, BERLIN, Germany, D-10243
FEATURES
Source
/organism="synthetic construct"
/db_xref="taxon:32630"
/focus
1. .1234
/organism="Bacillus amyloliquefaciens"
/db_xref="taxon:1390"
1. .864
/organism="Bacillus amyloliquefaciens"
/db_xref="taxon:1390"
1. .1234
/organism="Paenibacillus macerans"
/organism="taxon:44252"
469. .543
/function="endo-1,3-1,4-beta-glucanase"
/feature="Hybrid between bgla (Pos. 1 to 864) and bgalm
(Pos.in hybrid 865 to 1234, Pos. original 463 to 852)"
/codon_start=1
/transl_table=11
/label="HA107M"
/evidence="experimental"
/product="hybrid-endo-1,3-1,4-beta-glucanase"
/protein_id="CAA81093.1"
/db_xref="GI:398188"
/translation="MKRWNVNMTSLGEMRLALTSYKFKDCGENRSVOTGYGLYEV
KADGYSQGNMCTWRANNVSTSLGEMRLALTSYKFKDCGENRSVOTGYGLYEV
RMKPAKNTGIVSSFTYTGPTGTDEIDIEFLGKDTTKVQFNYYNGVGHKVIS
LGFDAKSGPHYAFDMQPGYIKWYVDGVLKHTATANIPSTPGKIMNMLNNGTGVDDWL
GSGNANPLYAEYDMVKYTSN"
544. .1234
/organism="Paenibacillus macerans"
/db_xref="taxon:44252"
BASE COUNT    369 a 222 c 285 g 358 t
ORIGIN

Query Match      5.7%; Score 42.2; DB 56; Length 1234;
Best Local Similarity 54.7%; Pred. No. 0.75;
Matches 111; Conservative 0; Mismatches 83; Indels 9; Gaps 1;

QY 146 cagatggaaggccctgggtagaagtggatattgaagttctcgcgaagaatccgggcagtt 205
||||| || || || || || || || || || || || || || || || || || || ||
Db 836 CGGAGGGGACTCCTGGGATGAGATTGATATCGAATTTCTAGGAAAGACACACAAAG 895
||||| || || || || || || || || || || || || || || || || || || ||

QY 206 tcagtcaccaatcattaccggaagccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 265
||||| || || || || || || || || || || || || || || || || || || ||
Db 896 TCAGTTTAACTATTATACCAATGGGTTGGCGGTCTAT-----GAAAGGTTATCT 946
||||| || || || || || || || || || || || || || || || || || || ||

QY 266 ctgttagcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 325
||||| || || || || || || || || || || || || || || || || || || ||
Db 947 CTCTTGGCTTTGATGATCATCAAGGCTTCCATACCTATGCTTTCGATTGCGACCCAGGT 1006
||||| || || || || || || || || || || || || || || || || || || ||

QY 326 acgtccgctggactgtgacggt 348
||||| || || || || || || || || || || || || || || || || || || ||
Db 1007 ATATTAATGGTATGACCGT 1029
||||| || || || || || || || || || || || || || || || || || || ||

Search completed: August 23, 2001, 13:37:21
Job time: 8284 sec

```


GenCore version 4.5
 Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 11:12:42 ; Search time 157.43 Seconds
 (without alignments)
 2967.406 Million cell updates/sec

Title: US-09-654-652A-4
 Perfect score: 744
 Sequence: 1 atggttagcgaagattt.....gagatgacgaacctgtctcg 744

Scoring table:
 IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : N_Geneseq_0601:*

- 1: /cgnl_9/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /cgnl_9/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /cgnl_9/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /cgnl_9/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /cgnl_9/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /cgnl_9/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /cgnl_9/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /cgnl_9/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /cgnl_9/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /cgnl_9/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /cgnl_9/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /cgnl_9/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /cgnl_9/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /cgnl_9/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /cgnl_9/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /cgnl_9/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /cgnl_9/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /cgnl_9/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /cgnl_9/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /cgnl_9/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /cgnl_9/gcgdata/geneseq/geneseq/NA2000.DAT:*
- 22: /cgnl_9/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47.4	6.4	927	20	AA02912 B. alkalophilus be
2	44.6	6.0	783	11	AA005167 Heat-stable endo-b
3	44.6	6.0	850	11	AA003519 Thermostable beta-
4	42.2	5.7	1240	11	AA005832 Beta-glucanase-H1
5	35.2	4.7	445	21	AA071502 Neocarcinostatin a
6	34	4.6	464	21	AA071503 Neocarcinostatin a
7	34	4.6	971	19	AAV29067 Lichenase coding s
8	34	4.6	2104	22	AA087120 Nucleotide sequenc
9	33.6	4.5	1459	17	AA029043 Oerskovia beta-1,3
10	33.4	4.5	556	21	AA096632 Fusarium venenatum
11	32.8	4.4	1604	19	AAV20465 Human c-ets oncoge

12	32.8	4.4	1680	21	AA07212 P. oleoverans PHA
13	32.8	4.4	3970	14	AAQ38219 NANBH virus strain
14	32.8	4.4	4500	12	AAQ10503 6kB fragment contg
15	32.8	4.4	9589	14	AAQ38218 NANBH virus strain
16	32.8	4.4	9711	22	AA086644 Nucleotide sequenc
17	32.8	4.4	9711	22	AA086937 Nucleotide sequenc
18	32.8	4.4	9711	22	AA023486 Infectious Hepatit
19	32.8	4.4	15872	18	AA068715 Streptomyces venez
20	32.8	4.4	15872	21	AA087283 S. venezuelae vep
21	32.6	4.4	1590	21	AA083394 Human 4F2 heavy ch
22	32.6	4.4	1863	21	AA083374 Human 4F2 heavy ch
23	32.6	4.4	1897	22	AA077279 Human CDNA encodin
24	32.4	4.4	10732	21	AA010594 Gene encoding a su
25	32.2	4.3	622	21	AA069121 Fusarium venenatum
26	32.2	4.3	3401	11	AA069094 Sequence encoding
27	32	4.3	1247	21	AA013476 Aspergillus oryzae
28	31.8	4.3	1063	10	AA095000 Bacillus subtilis
29	31.8	4.3	72750	21	AA081468 N. meningitidis pa
30	31.8	4.3	34980	21	AA021544 Neisseria meningit
31	31.6	4.2	453	21	AA067724 Eucalyptus grandis
32	31.6	4.2	1473	20	AA080867 Degenerate DNA seq
33	31.6	4.2	6564	22	AA031637 Mycobacterium tube
34	31.4	4.2	691	21	AA011284 Aspergillus niger
35	31.4	4.2	987	18	AA089157 Oerskovia xanthine
36	31.4	4.2	1177	18	AA089155 Oerskovia xanthine
37	31.4	4.2	1516	18	AA089156 Oerskovia xanthine
38	31.2	4.2	601	21	AA011435 Aspergillus niger
39	31.2	4.2	667	15	AA058486 HCV peptide C14-8.
40	31.2	4.2	667	15	AA058823 NANBH virus gene f
41	31.2	4.2	3220	14	AA052535 Encodes Subtilisin
42	30.8	4.1	490	19	AA038809 Homo sapiens CESP
43	30.6	4.1	1347	22	AA060981 P. putida KT2440-a
44	30.6	4.1	5021	21	AA034703 Human tolloid-like
45	30.4	4.1	44377	18	AA078508 Platenolide syntha

ALIGNMENTS

RESULT 1

AA02912

ID AA02912 standard; DNA; 927 BP.

XX

AC AA02912;

XX

DT 19-MAY-1999 (first entry)

XX

DE B. alkalophilus beta-glucanase coding region DNA.

XX

XX Beta-glucanase; enzyme; glucan; removal; lichenan; membrane;

KW equipment; food industry; brewing; ds.

XX

OS Bacillus alkalophilus.

XX

XX

PN DEL9732751-AL.

XX

PD 04-FEB-1999.

XX

PF 30-JUL-1997; 97DE-1032751.

XX

PR 30-JUL-1997; 97DE-1032751.

XX

XX (HENK) HENKEL KGAA.

PA

XX

XX Hillen W, Maurer K;

PI

XX

XX WPI; 1999-122161/11.

DR

XX

XX P-PSDB; AAW93001.

DR

XX

XX New Bacillus alkalophilus beta-glucanase protein and gene - useful

PT for removing glucan and/or lichenan from membranes in the brewing

PT industry

XX

DR P-PSDB; AAR03775.

XX Thermotable beta-glucanase production -
PT using Bacillus subtilis transformed with gene from Bacillus
PT macerans.

XX Fig 1; ; 9pp; German.

XX The gene encoding Bacillus macerans thermostable
CC beta-glucanase is expressed in Bacillus subtilis. The
CC enzyme is useful for lowering the viscosity of brewing
CC mashes and in the production of feedstuff.

XX Sequence 850 BP; 250 A; 160 C; 196 G; 244 T; 0 other;
SQ

Query Match 6.0%; Score 44.6; DB 11; Length 850;

Best Local Similarity 52.9%; Pred. No. 0.00043;
Matches 154; Conservative 0; Mismatches 119; Indels 18; Gaps 2;

QY 58 tacggttaagttaagcccgatgaagatggcagccgcatcggaacagtcagttccatg 117

DB 373 tacggcctgacgaggtcagtcagtcagccagcccaaaatacaggaattgtctcatcttt 432

QY 118 ttctctaccagaatggttcgaaatccgcatggaagccctgggtagagtgatatt 177

DB 433 ttacgta-----tacagacctgctcatggcacaacatgggatgaatagatc 483

QY 178 gaagtctcgcaagaatccggcgagtttccagtcacacatcattaccggaagccgcg 237

DB 484 gaattctagaaagacacagcaaaagtcagtttaactattatatacgaatgggttgcg 543

QY 238 gcacaaagactagcgaagcaccatgctgttagcccgcccgcatcaggtttccac 297

DB 544 ggtcat-----gaaaggttattctctctgtggttgatcatcaagggcttccat 594

QY 298 acctacggttcgaatggaactcgaattacgtccgctggactgttgacggt 348

DB 595 acctatgcttcgattggcagcaggtatataaaatggatgtagacggt 645

RESULT 4

AAQ05832
ID AAQ05832 standard; DNA; 1240 BP.

XX AC AAQ05832;

XX 09-JAN-1991 (first entry)

XX DE Beta-glucanase-H1 gene.

XX Hybrid pre-beta-glucanase; glucans; beer; animal feed; poultry; ss.

XX Bacillus amyloliquefaciens, Bacillus macerans.

XX Key Location/Qualifiers

FT CDS 469..1188

FT /*tag= a

FT /label=Hybrid pre-glucanase

FT misc_RNA 469..864

FT /*tag= b

FT /label=N-terminal of B.amyloliquefaciens beta-

FT glucanase

FT mISC_RNA 865..1185

FT /*tag= c

FT /label=C-terminal half of B.macerans beta-

FT glucanase

XX WO9009436-A.

XX PN 23-AUG-1990.

XX PD 16-FEB-1990; 90WO-DK00044.

XX PF

XX

PR 04-AUG-1989; 89DK-0003848.

PR 16-FEB-1989; 89DD-0325800.

XX (CARL-) CARLSBERG A/S.

PA (DEAK) AKAD WISSENSCHAFT DDR.

XX

PI Borriass R, Hofemeister J, Thomsen KK, Olsen O, Vonwettstein D;

XX

XX WPI; 1990-275129/36.

DR P-PSDB; AAR06621.

XX

PT New thermostable (1,3-1,4)-beta-glucanase - prepd. using hybrid

PT gene obt'd. using Bacillus amyloliquefaciens and B.macerans genes

XX

XX Disclosure; page 26; 84pp; English.

XX

CC This beta-glucanase-H1 gene encodes a hybrid pre-beta-glucanase

CC comprising a signal peptide and the amino-terminal of the B.

CC amyloliquefaciens beta-glucanase and the carboxyl-terminal half of

CC the B.macerans beta-glucanase. The hybrid (1,3-1,4)-beta-gluc-

CC anase-H1 encoded is thermostable and hydrolyses beta-glycosidic

CC linkages in (1,3-1,4)-beta-glucans. Reducing sugars are obtd. at

CC high temps. and thus this enzyme can be used in the mfr. of food

CC prods., esp. beer and animal feed (e.g. for feeding poultry).

XX

XX See also AAQ05833.

SQ

Sequence 1240 BP; 370 A; 223 C; 287 G; 360 T; 0 other;

XX

Query Match 5.7%; Score 42.2; DB 11; Length 1240;

Best Local Similarity 54.7%; Pred. No. 0.0029;

Matches 111; Conservative 0; Mismatches 83; Indels 9; Gaps 1;

QY

146 ccattgaagccctggtagaagtcattgaatttcgcaagaatccggcagtt 205

DB 836 cggaggggactccttgggtagagattgatatcggaattcttaggaagacacacaaag 895

QY 206 tccagtcacacatcattaccgggtaagccgcccgcacacaaagactagcgaagacccatg 265

DB 896 tccagtttaactattataccaatgggttggtggtcat-----gaaaaggttatct 946

QY 266 ctgttagcccgcccgatcaggctttccacacctacgctcgaatggactccgaatt 325

DB 947 ctcttgctttgatgcataaagggttccataacctatgcttcttgatggcagccaggt 1006

QY

326 acgtccgctgactgttgacggt 348

DB

1007 atataaatggtagtagcgggt 1029

XX

RESULT 5

AAAT71502

ID AAA71502 standard; DNA; 445 BP.

XX

AC AAA71502;

XX

DT 11-DEC-2000 (first entry)

XX

DE Neocarzinostatin apoprotein DNA fragment.

XX

XX Neocarzinostatin; NCS; apoprotein; apoNCS; chemotherapy;

XX acute leukemia; bladder cancer; pancreatic cancer ds.

XX OS Synthetic.

XX

XX JP2000175687-A.

XX

XX 27-JUN-2000.

XX

XX 16-DEC-1998; 98JP-0358029.

XX

XX 16-DEC-1998; 98JP-0358029.

XX

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 11:12:42 ; Search time 1874.7 Seconds
(without alignments)
3751.493 Million cell updates/sec

Title: US-09-654-652A-4
Perfect score: 744
Sequence: 1 atggttagcgcaaggattt.....gagatgacgaacctgctccg 744

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1:	gb_estl1:*
2:	gb_est2:*
3:	gb_est3:*
4:	gb_est4:*
5:	gb_est5:*
6:	gb_est6:*
7:	gb_est7:*
8:	gb_est8:*
9:	gb_est9:*
10:	gb_est10:*
11:	gb_est11:*
12:	gb_est12:*
13:	gb_est13:*
14:	gb_est14:*
15:	gb_est15:*
16:	gb_est16:*
17:	gb_est17:*
18:	gb_est18:*
19:	gb_est19:*
20:	gb_est20:*
21:	gb_est21:*
22:	gb_est22:*
23:	gb_est23:*
24:	gb_est24:*
25:	gb_est33:*
26:	gb_est34:*
27:	gb_est35:*
28:	gb_est36:*
29:	gb_est37:*
30:	gb_est38:*
31:	gb_est39:*
32:	gb_est40:*
33:	em_estba:*
34:	em_estfun:*
35:	em_esthum1:*
36:	em_esthum2:*
37:	em_esthum3:*
38:	em_esthum4:*
39:	em_esthum5:*
40:	em_esthum6:*
41:	em_esthum7:*
42:	em_esthum8:*
43:	em_esthum9:*
44:	em_esthum10:*
45:	em_esthum11:*
46:	em_esthum12:*
47:	em_esthum13:*
48:	em_esthum14:*
49:	em_esthum15:*
50:	em_esthum16:*
51:	em_esthum17:*
52:	em_esthum18:*
53:	em_esthum19:*
54:	em_esthum20:*
55:	em_esthum21:*
56:	em_esthum22:*
57:	em_esthum23:*
58:	em_esthum24:*
59:	em_esthum25:*
60:	em_esthum26:*
61:	em_esthum27:*
62:	em_esthum28:*
63:	em_estin1:*
64:	em_estin2:*
65:	em_estin3:*
66:	em_estin4:*
67:	em_estin5:*
68:	em_estomi:*
69:	em_estom2:*
70:	em_estov1:*
71:	em_estov2:*
72:	em_estpl1:*
73:	em_estpl2:*
74:	em_estpl3:*
75:	em_estpl4:*
76:	em_estpl5:*
77:	em_estpl6:*
78:	em_estpl7:*
79:	em_estpl8:*
80:	em_estpl9:*
81:	em_estpl10:*
82:	em_estro1:*
83:	em_estro2:*
84:	em_estro3:*
85:	em_estro4:*
86:	em_estro5:*
87:	em_estro6:*
88:	em_estro7:*
89:	em_estro8:*
90:	em_estro9:*
91:	em_estro10:*
92:	em_estro11:*
93:	em_estro12:*
94:	em_estro13:*
95:	em_estro14:*
96:	em_estro15:*
97:	em_estro16:*
98:	em_estro17:*
99:	em_estro18:*
100:	em_estro19:*
101:	em_estro20:*
102:	gb_est25:*
103:	gb_est26:*
104:	gb_est27:*
105:	gb_est28:*
106:	gb_est29:*
107:	gb_est30:*
108:	gb_est31:*
109:	gb_est32:*
110:	gb_est41:*
111:	gb_est42:*
112:	gb_est43:*
113:	gb_est44:*
114:	gb_est45:*
115:	gb_est46:*
116:	gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

source

JOURNAL
COMMENT
in rice blast fungus, Magnaporthe grisea
Unpublished (1998)
Contact: Dean, R.A.

ALIGNMENTS

CHAZ, D.W., CROOKER, S., GONZALEZ, J., HAN, P.S., HSLA, C.C., KANG, Y., LAZO, G.R., MILLER, R., RAUSCH, C.J., ROSS, K., SEATON, C.L. and TONG, J.C.

ORIGIN

in Bonaldo et al (1996) Genome Research 6: 791-806."

```

BASE COUNT 153 a 240 c 222 g 132 t
ORIGIN

Query Match
Best Local Similarity 4.9%; Score 36.8; DB 170; Length 747;
Matches 107; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 488 tcaactgggtcaaggtttataagttacacgcccagggccaggaagcgcagcgacttta 547
Db 420 TCNAGTATGACTCGGTGCCACAGACTGCCCCGGCCGACCTCAATGCGACGACGAGCGGT 479

QY 548 cgttgactggaccgacaattttgacacgtttgatgctcccgctgggcaaggtgact 607
Db 480 TCTGTGTGAGGGCGCGCAAGATTACACCTTTACTGAGAGCGACCGCAGACATCACT 539

QY 608 gacatttgacggttaaccgtgtgacactcaacgacaagaacatctactccagagatgca 667
Db 540 GGGGCGCTCGCGGCGTGCATCGTATTGAGAGCACTGACGCTGTTCACTGACATCGGCA 599

QY 668 tgttgatctcgccctcacccgcaaaaggtcaggaagcttcaac 711
Db 600 AGCCACCGCCCACTGACCGCGGCCCAAGAAGGTGATCATC 643

RESULT 7
LOCUS BE359676 604 bp mRNA EST 20-JUL-2000
DEFINITION DGL_56_H08.b2_A002 Dark Grown 1 (DGL) Sorghum bicolor CDNA, mRNA
ACCESSION BE359676
VERSION BE359676.1 GI:9301233
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 604)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.
TITLE An EST database from Sorghum: dark-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 536
POLYA-No.
Location/Qualifiers
1..604
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DGL)"
/note="Organ: 5-day-old dark-grown seedlings; Vector:
Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
made from polyA RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."

BASE COUNT 128 a 208 c 167 g 101 t
ORIGIN

Query Match
Best Local Similarity 4.9%; Score 36.4; DB 156; Length 604;
Matches 91; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 166 gaagtggatattgaagttctcggcaagaatccgggagttcttcagttcccaatcattacc 225
Db 421 GAGATCGACTTCGAGTTCTCGCAACGTCCTCGGGGAGCCCTACACGCTGCACACCAAC 480

QY 226 gtaaggccggcgacaaagactagcgaagcaccatctgttttagcccccgcgcgat 285
Db 481 GTGTTCACCGGGCGGCGAGCCGAGCGAGCGACAGTTCGGCCTCTGTTCGACCCACC 540

QY 286 caggtttccacactcaggtctcgaatgactcgaatagctccgctggactgttgc 345
Db 541 AAGGACTTCCACACCTACTCCGTGGTGTGGAACCCGACGACGTGATCTTCGCCGTGGAC 600

QY 346 gg 347
Db 601 GG 602

RESULT 8
LOCUS BG488141 608 bp mRNA EST 27-MAR-2001
DEFINITION RHIZ2_60_A07.b1_A003 Rhizome2 (RHIZ2) Sorghum propinquum CDNA, mRNA
ACCESSION BG488141
VERSION BG488141.1 GI:13469376
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 608)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
,L.H.
TITLE An EST database from Sorghum: Sorghum propinquum rhizomes
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Seq primer: JEN REV
High quality sequence stop: 535
POLYA-No.
Location/Qualifiers
1..608
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RHIZ2)"
/note="Organ: Rhizomes; Vector: pBluescript II from Lambda
Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."

BASE COUNT 129 a 208 c 169 g 102 t
ORIGIN

Query Match
Best Local Similarity 4.9%; Score 36.4; DB 154; Length 608;
Matches 91; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 166 gaagtggatattgaagttctcggcaagaatccgggagttcttcagttcccaatcattacc 225
Db 425 GAGATCGACTTCGAGTTCTCGCAACGTCCTCGGGGAGCCCTACACGCTGCACACCAAC 484

QY 226 gtaaggccggcgacaaagactagcgaagcaccatctgttttagcccccgcgcgat 285
Db 485 GTGTTCACCGGGCGGCGAGCCGAGCGAGCAGTTCGGCCTCTGTTCACCCACC 544

QY 286 caggtttccacactcaggtctcgaatgactcgaatagctccgctggactgttgc 345
Db 545 AAGGACTTCCACACCTACTCCGTGGTGTGGAACCCGACGACGTGATCTTCGCCGTGGAC 604

```

DD GCATCAACCGGA--AACATCCGGGCACTGTCTCCTGCCTTCTACCTTTCTCCGGTGA 196

[illegible]

4.88; Score 36; DB 155; Length 599;

Qy	507	taagtatacccgccgagcgccgaagcgccagcgaacttaoacttgactgaccgacaa	566
Db	465	CGCGATCGGGGTCCGACAAGCGTGGTGCCTCGAGGACTATGTCTATCGAGGTCGACAAGCG	406
Qy	567	ttttgacacgtttgatggtctccgcgtggggcaagggtgactgacattttgacggtaacccg	626
Db	405	TCACGGTGGTGTTCGAGGACTACGCCATCGGGGTCGATAAGCGTGGTCTGTCTGAGGACTA	346
Qy	627	tgtcgacctcaccgcaagaacatcttactctcagagatggcatgttgatctgcgcctcac	686
Db	345	CGTGATCCCGATCGACAAGCGCCACAACACTCCGTCGAGGACTATGTGATCGAGGTCGACAA	286
Qy	687	ccgcaaa	693
Db	285	GGCTAA	279

QY	067	ccgcaca	895
Db	285	GGCCTAA	279
RESULT	14		
LOCUS	BE888160/c		
DEFINITION	BE888160 sequence.		
ACCESSION	BE888160		
VERSION	BE888160.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 640)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs@mail.nih.gov Tissue Procurement: ATCC CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LIAW9732 row: n column: 18 High quality sequence stop: 637.		

```

source
i. 040
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3913169"
./clone_lib="NIH_MGC_71"
./tissue_type="leiomyosarcoma"
/lab_host="PH10B (phage-resistant)"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site.1: No
Site.2: Sall; Cloned unidirectionally. Primer: Oligo
Average insert size 2.1 kb. "
BASE COUNT      147 a      196 c      169 g      127 t      1 others
ORIGIN

Query Match      4.8%; Score 35.4; DB 141; Length 640;
Best Local Similarity 53.2%; Pred. No. 6;
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps

```

Db	515	GAGCCCTGGATGTTAGAAAGCGACAGAGGTCCTCAATATGCACAGTGGCCCTTTTGGCAA	456
Qy	388	ttgacaggtacacagggaactccgttttaacctttgtctgcgtgagagtcgagcttggggt	447
Db	455	AATGTCGGCAGACAGGCGGTCGAGTTTGTCTCCATAACCTGGGGGTATCCTTGGGGG	396
Qy	448	qgccagttccatgaatcaaaq	468

```
Db 395 GCAGGTATTCTGTAGAG 375
||| || | || |||
||| || | || |||

RESULT 15
BE208885 669 bp mRNA EST 28-JUN-2000
GF-FV-P3C5 Marsh grapefruit young flavedo Citrus x paradisi cdna,
mRNA sequence.
ACCESSION BE208885
VERSION BE208885.1 GI:8789137
KEYWORDS EST.
SOURCE Citrus x paradisi.
ORGANISM Citrus x paradisi
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Sapindales; Rutaceae; Citrus.
REFERENCE 1 (bases 1 to 669)
AUTHORS McCollum,T.G., Doostdar,H., Shatters,R. and Mayer,R.T.
TITLE Marsh grapefruit young flavedo EST
JOURNAL Unpublished (2000)
COMMENT Contact: McCollum, T.G.
US Horticultural Research Laboratory
USDA, ARS
2001 S. Rock Road, Ft. Pierce, FL 34945, USA
Tel: 561-462-5836
Fax: 561-462-5986
Email: gmcollum@ushrl.ars.usda.gov
Plate: 3 row: C column: 5.
FEATURES
    source
        1..669
            Location/Qualifiers
                /organism="Citrus x paradisi"
                /cultivar="Marsh"
                /db_xref="taxon:37656"
                /clone_lib="Marsh grapefruit young flavedo"
                /tissue_type="Flavored"
                /dev_stage="2-5 gram fruit"
                /note="Organ: Fruit"
BASE COUNT 163 a 138 c 187 g 181 t
ORIGIN

Query Match 4.8%; Score 35.4; DB 164; Length 669;
Best Local Similarity 51.6%; Pred. No. 6.1;
Matches 81; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 207 ccagtcacaacatcattaccggttaaggccggcgacacaaagactagcgaaagaccatgc 266
|| |||| | | |||| | | |||| | | |||| | | |||| | | |||| | | ||||
Db 41 CCGTACCAAGCAAACTGCACGTAGTCACCGGAGAAAGGCTCCTAGGAGCAACTTGC 100
| | |||| | | |||| | | |||| | | |||| | | |||| | | |||| | | ||||
QY 267 tgttagcccccgcgcgacatcaggctttccacacctacggtctcgaatgcactccgaatta 326
| | |||| | | |||| | | |||| | | |||| | | |||| | | |||| | | ||||
Db 101 TACCAAGGCCGCCCGTAGCTGCGCCACACCGGTGTTAAGAGCCCTCATCGTTA 160
| | |||| | | |||| | | |||| | | |||| | | |||| | | |||| | | ||||
QY 327 cgtccgctgactgttgacggttcaggagtcgcgcaag 363
| | |||| | | |||| | | |||| | | |||| | | |||| | | |||| | | ||||
Db 161 CGGTCTGGACCGTCGCCCTTCGAGAAATCGCAAG 197
| | |||| | | |||| | | |||| | | |||| | | |||| | | |||| | | ||||
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 12:14:13 ; Search time 2906.88 Seconds
(without alignments)
3958.886 Million cell updates/sec

Title: US-09-654-652A-4
Perfect score: 744
Sequence: 1 atggttagcgcaaggattt.....gagatgacgaacctgctcg 744

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 0
Total number of hits satisfying chosen parameters: 2686314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:

1: gb_bal:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_in4:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_bal:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_inv3:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vil:*
59: gb_vil2:*
60: gb_vil3:*
61: gb_vil4:*
62: gb_vil5:*
63: gb_vil6:*
64: gb_vil7:*
65: gb_vil8:*
66: gb_vil9:*
67: gb_vil10:*
68: gb_vil11:*
69: gb_vil12:*
70: gb_vil13:*
71: gb_vil14:*
72: gb_vil15:*
73: gb_vil16:*
74: gb_vil17:*
75: gb_vil18:*
76: gb_vil19:*
77: gb_vil20:*
78: gb_vil21:*
79: gb_vil22:*
80: gb_vil23:*
81: gb_vil24:*
82: gb_vil25:*
83: gb_vil26:*
84: gb_vil27:*
85: gb_vil28:*
86: gb_vil29:*
87: gb_vil30:*
88: gb_vil31:*
89: gb_vil32:*
90: gb_vil33:*
91: gb_vil34:*
92: gb_vil35:*
93: gb_vil36:*
94: gb_vil37:*
95: gb_vil38:*
96: gb_vil39:*
97: gb_vil40:*
98: gb_vil41:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	741	99.6	1426	3	FIBGLUC
2	20	2.7	1289	12	AB009683
3	20	2.7	8435	5	AF236106
4	20	2.7	70841	4	AC004287
5	20	2.7	93588	64	AC017187
6	20	2.7	268219	5	AE003806
7	19	2.6	1944	97	HSU45983
8	19	2.6	2608	97	HSU62556

M33676 F.succinog
AB009683 Raphanus
AF236106 Drosophil
AC004287 Drosophil
AC017187 Drosophil
AE003806 Drosophil
U45983 Homo sapien
U62556 Human chemo

```
c 9          2.6 2608 97 H5U62556      UG2556 Human chemo
c 10         2.6 86172 91 AY016370 Homo sapi
c 11         2.6 86172 91 AY016370 Homo sapi
c 12         2.6 98861 89 AL138961 Human DNA
c 13         2.6 147009 76 AC084179 Homo sapi
c 14         2.6 151174 64 AC016189 Homo sapi
c 15         2.6 159852 70 AC026309 Homo sapi
c 16         2.6 159852 70 AC026309 Homo sapi
c 17         2.6 166826 80 AL356535 Homo sapi
c 18         2.6 167506 69 AC025679 Homo sapi
c 19         2.6 178563 83 AP002419 Homo sapi
c 20         2.6 194014 81 AL583784 Homo sapi
c 21         2.6 213524 61 AC010735 Homo sapi
c 22         2.6 220055 74 AC069499 Homo sapi
c 23         2.4 86 9 AR045081
c 24         2.4 86 10 IL8613
c 25         2.4 86 10 IL8613
c 26         2.4 702 54 G51819
c 27         2.4 1213 95 RATGSTRPS
c 28         2.4 1333 94 AF156988
c 29         2.4 1352 94 AF156988
c 30         2.4 1381 1 AF037600
c 31         2.4 1532 3 STMSALO
c 32         2.4 3602 14 CPSUCPSY2
c 33         2.4 5591 1 AF027499
c 34         2.4 7425 2 AX024365
c 35         2.4 7425 9 AX024258
c 36         2.4 7805 65 AC018084
c 37         2.4 10145 1 AE000839
c 38         2.4 13416 1 AE000117
c 39         2.4 34074 3 SCF81
c 40         2.4 34766 3 SC121
c 41         2.4 37931 3 SCD10
c 42         2.4 41565 6 CF57A10B
c 43         2.4 47713 2 AX024320
c 44         2.4 47713 9 AX024213
c 45         2.4 47887 63 AC015109

ALIGNMENTS

RESULT 1
FIBGLUC      1426 bp      DNA      BCT      26-APR-1993
LOCUS        F. succinogenes 1,3-1,4-beta-D-glucan 4-glucanohydrolase gene,
DEFINITION   complete cds.
ACCESSION    M33676 M3311
VERSION      M33676.1 GI:148575
KEYWORDS     1,4-beta-D-glucan 4-glucanohydrolase; mixed-linkage beta-glucanase.
SOURCE       F. succinogenes (strain S85) DNA, clone PJ15.
ORGANISM     Fibrobacter succinogenes
Bacteria; Fibrobacter/Acidobacteria group; Fibrobacter group;
Fibrobacter.
REFERENCE    1 (bases 1 to 1426)
AUTHORS      Teather, R.M. and Erfile, J.D.
TITLE        DNA sequence of a Fibrobacter succinogenes mixed-linkage
            beta-glucanase (1,3-1,4-beta-D-glucan 4-glucanohydrolase) gene
JOURNAL      J. Bacteriol. 172, 3837-3841 (1990)
MEDLINE      90299807
COMMENT      Draft entry and computer-readable sequence for [1] kindly submitted
            by R.M. Teather, 11-APR-1990.
FEATURES     source
            Location/Qualifiers
            1..1426
            /organism="Fibrobacter succinogenes"
            /db_xref="taxon:833"
            62..66
            -35_signal
            85..90
            RBS      132..137
            /note="ribosome binding site"
            145..1194
            CDS      /note="1,3-1,4-beta-D-glucan 4-glucanohydrolase precursor
            (EC 3.2.1.73)"
```

```
/codon_start=1
/transl_table=11
/protein_id="AAA24896.1"
/db_xref="GI:148576"
/translation="MNKKTKAVKSAALAVAAAAAATTNVSADKPSGAEYLTLEVOYK
KFEARMKMAASGVSSMFLYONGSEADGRPWVEVDIEVLGKNPQSNITGKAG
AKTSEKHHAIVSPADQAFHYTGLEWTPNYVWTVDGVEGKSDFTLQWTFDTGLR
FNLSSESAAWVGQFDESKLPLOF INNVKVKYTPGQEGGSDFTLQWTFDTGLR
SRWCKGWTEDGNRDLTKNYSRDGMLIALPKQGESFNGQVPRDEPAPOSSSS
APASSSVPPASSVPASSSSAFVPPSSSSATNAIHGMRITTPAVAKHRNLVNAKAG
VNPNGHKRYRVNFEH"
sig_peptide 145..225
            /note="1,3-1,4-beta-D-glucan 4-glucanohydrolase signal
            peptide"
mat_peptide 226..1191
            /note="1,3-1,4-beta-D-glucan 4-glucanohydrolase"
BASE COUNT 371 a 346 c 335 g 374 t
ORIGIN
Query Match 99.6%; Score 741; DB 3; Length 1426;
Best Local Similarity 100.0%; Pred. No. 0; Indels 0; Gaps 0;
Matches 741; Conservative 0; Mismatches 0;
Qy 4 gtagcgcaaaagatttagcggtgcgaactctcacagttagaagaagttcagtcaggt 63
|||||
Db 217 GTTAGCGCAAGAGATTATTAGCGGTCCGACCTACACGTTAGAGAGAGTTTCCAGTACGGT 276
Qy 64 agtttgagccggtatgaagatggcagcgcatcggaacagtcagttccatgtctctc 123
|||||
Db 277 AAGTTTGAAGCCGCTATGAAGATGGCAGCGCATCGGGAACAGTACGTTCCATGTTCTCTC 336
Qy 124 taccagaatgttcgaatacgcgaatgaagagccctgggtagaagtagatgaattgaagt 183
|||||
Db 337 TACCAGATGGTTCCGAATTCGCCGATGGAAGGCCCTGGGTAGAAGTGGATATTGAAGTT 396
Qy 184 ctgcgcaagaatccggcgagttctccagttcccaacatcataccggttaagccggcgcaaca 243
|||||
Db 397 CTCGGCAGAGATCCGGGCGAGTTTCCAGTCCACATCATATTACCGGTAAAGCCGGCGCACAA 456
Qy 244 agactagcgaagaagcaccatgctgttagcccgccgcatcaggctttccacacctac 303
|||||
Db 457 AAGACTAGCGAAAGACCATGCTGTAGCCCGCCGCGATCAGGCTTTCACACCTTAC 516
Qy 304 ggtctcgaatggactccgaataatcagtcgctggaactgtgacggttcaggaagtcgcaag 363
|||||
Db 517 GGTCTCGAATGGACTCCGAAATTTACGTCCGCTGGAGCTGTGACGGTTCAGGAAGTCCGCAAG 576
Qy 364 acggaagtgccaggtttccaaactgcaggtacacagggactccgttttaaccttgg 423
|||||
Db 577 ACGGAAGTGGCCAGGTTTCCAACTTGACAGGTACACAGGAGTCCGTTTAACTTTTGG 636
Qy 424 tcgtctgagatgctggcttgggtggccagttcgatgaatcaaaagcttcctcgtttccag 483
|||||
Db 637 TCGTCTGAGAGTGGCGCTTGGTTCGCCAGTTTCGATGAATCAAAAGCTTTCGCTTTTCCAG 696
Qy 484 ttcataactgggttcgaagtttataagtatatacgccggccagggcgagggcgagcgac 543
|||||
Db 697 TTCATCAACTGGGTCAAGGTTTATAAGTATACGCCGGCCAGGGCGAAGGGCGGACAGC 756
Qy 544 tttaagcttgactgagcagcaattttgacacgtttgagctcccgctggggcgaaggt 603
|||||
Db 757 TTTACGCTTGACTGGACCGACAAATTTTACACAGTTTGATGCTCCCGTGGGGCAAGGTT 816
Qy 604 gactggacatttgacggttaaccgtgtgcacctcaccgcaagaatactactctccagagat 663
|||||
Db 817 GACTGGACATTTGACGGTAAACCGTGTGCACCTCACCCGACAAAGAACATCTACTCCAGAGAT 876
Qy 664 ggcattgtgactcctgcctcaccgcgaagaaagtcaggaagaaagcttcaacggcaggttccg 723
|||||
Db 877 GGCATGTTGATCTCTCGCCCTCACCGCAAGGTCAGAAAGCTTCAACGGCGCAGGTTCCG 936
Qy 724 agagatgacgaacctgctccg 744
```



```
|||||
Db 937 AGATGACGAACCTGCTCCG 957

RESULT 2
AB009683 1289 bp DNA PLN 25-JUL-1998
LOCUS Raphanus sativus SLG(S8) gene for S glycoprotein, partial cds.
DEFINITION AB009683
ACCESSION AB009683
VERSION SLG(S8); 1 GI:3327851
KEYWORDS Raphanus sativus DNA.
SOURCE Raphanus sativus
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Raphanus.
REFERENCE 1 (bases 1 to 1289)
AUTHORS Kusaba, M. and Nishio, T.
TITLE Direct Submission
JOURNAL Submitted (11-DEC-1997) to the DDBJ/EMBL/GenBank databases. Makoto
Kusaba, National Institute of Agrobiological Resources, Institute
of Radiation Breeding; 2425 Kamimurata, Ohmiya, Ibaraki 319-22,
Japan (E-mail:Kusaba@irb.affrc.go.jp, Tel.81-2955-2-1138,
Fax:81-2955-3-1075)
2 (sites)
REFERENCE Sakamoto, K., Kusaba, M. and Nishio, T.
AUTHORS Polymorphism of the S-locus glycoprotein gene (SLG) and the S-locus
TITLE related gene (SLR1) in Raphanus sativus L. and self-incompatible
JOURNAL ornamental plants in the Brassicaceae
MEDLINE Mol. Gen. Genet. 258 (4), 397-403 (1998)
FEATURES 98311079 Location/Qualifiers
source 1..1289
/organism="Raphanus sativus"
/db_xref="taxon:3726"
<1..>1289
/gene="SLG(S8)"
/feature="PCR product"
/product="S glycoprotein"
1..1289
/gene="SLG(S8)"
<1..>1289
/gene="SLG(S8)"
/codon_start=2
/product="S glycoprotein"
/protein_id="BAA31730.1"
/db_xref="GI:3327852"
/translation="DASYTLFLLVFFVLLFGPAHSINTLSSTESLTSSNRTLLSP
GNVFEGLFPKSSSRWYLGWTKLPDRYVWVNRDPLNSISIGTLKISNNLVLL
DHSKSWNTLNTRGERSPVVAELLANGNFVTRYNNNDSEFLWQDFDPTDLLP
EMKGLDLKGLNRLTSWKNDDPSSEISYQLDQRTGTFEFLKDGFRGYNRPW
NGVRNGIPEDQNLVWYNFTDSEVAVTFLTDKNTYSRLIISNDEYLAFLTLP
ASWDNLFTSPEPCDVYMGCPYACDVNTSPVCNIOGKPLNVQVQDLRDGSG
GCIRTELSGSGDGFIRMKNLPLDTMTATVDRSIGVKCEKRLSDCCTAFANADV
RNGGTGCVIWTGKLDIRNFYFDGQDLVRLAAADLV"
BASE COUNT 349 a 267 c 313 g 360 t
ORIGIN
Query Match 2.7%; Score 20; DB 12; Length 1289;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 637 accgaagaacatctactc 656
|||||
Db 758 ACCGACAGACATCTACTC 777

RESULT 3
AF236106 8435 bp mRNA INV 01-MAR-2000
LOCUS Drosophila melanogaster receptor protein tyrosine kinase ALK splice
DEFINITION variant A mRNA, complete cds.
ACCESSION AF236106
VERSION AF236106.1 GI:7110574
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 8435)
AUTHORS Palmer, R.H., Scully, A. Jr., Edeen, P., Thomas, J., McKeown, M. Jr. and
Hunter, T.
TITLE Identification of a novel Drosophila melanogaster RTK: dALK
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 8435)
AUTHORS Palmer, R.H., Scully, A. Jr., Edeen, P., Thomas, J., McKeown, M. Jr. and
Hunter, T.
TITLE Direct Submission
JOURNAL Submitted (17-FEB-2000) UCMP, Umea University, Building 6L, Umea
University, Umea S-901 87, Sweden
FEATURES 1..8435 Location/Qualifiers
source /organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2"
/map="53C-D"
378..5483
/codon_start=1
/product="receptor protein tyrosine kinase ALK splice
variant A"
/protein_id="AAF36990.1"
/db_xref="GI:7110575"
/translation="MPQLRVVLAISLLVDPSPFAQRPLINHTFTSSLPMSQHPSP
NSPSSAGGKQLRNLRLHEQLPLPQTLVATQYEDYEAAPSSGRDNKRRLAQMAG
RPGSGNGRGRLMESLRKELMNPVSGPGGSGGSGYPSYSATSSIRGDLGGGIG
SADRYGALYDOLRQPSGMAATAGPLGGYSTREPPLVAGVDQDQPSRNSKNSISELY
KLKALNQADEPAGGYTHGNEFGDQPPSSADPLOEIKDRIDTSPNANTYAPHTATP
SSEYELGVKCNFETPCSWTNGNSDGFVITGTLSKLNKLGLLPGPADSIDDN
GHFLYARVNPSSRLNLTSPFSTTEKCEFLVEMHSDMSHGLSRVYVELLHTAES
WVPAELIGDNVRQWTRVLRVSRDFRIVFVDPDLRVGQKHVALDRLMVNCFP
EDTKSEKTSQVKTSSKVPVCIHLPICRIDI TRDCDEADEQSCDKIPYGRCDFE
EDMGWRDCKTTLTWSRHTGSSPTDHTGPDGHTQHLQNTSGYMLVMNQHMNN
SEKNSIIGFASNAIMVSKTNPVPSVHGNPDSRYRSCVVFETHOFGKNPQNSISLV
VEMKENITTLWSTKNGSDWMAEYVLPNITSYIYLOFARMGWRYSIDVAVDD
FSLSPFCFLNIPEDHLGGYVWDRNLKSPYTKDFEYTNILEYTCDDTRGIPGSPQ
AQCEAAYREQNKTHLVREHVVEDQSSYKGMQKVPHEGHYTI IAKGSGGLGSGV
GSSRGSAVAAILLEHKEELYFLVGOGENACIKSMGLKACGCTDHDLDLAQYSFR
SKDMVKNIV IENGAGGGGSGYVFLLNQAKNEAPLLVAGGGGIGIGYIDEDQFH
GQAKPLQAPESQINGEPUNKKTAGPGGWRAKEDQALSPTYGAAALQGGRGHSCY
VELADNGTSVRRHGQGGGCGGCGCTGGGGYAGDVLYLTENSGEGSSYISPSRS
LRBSITHAGASSGPAIIIPAIEGCCDYRCVADLDFRSKVRICPDGWSLKRNDH
TACEIREEAKSGSFQVLSLIMISLAVLFICIAALIFMLYNRYORKKSRKMLVE
ODQLTRLRNNIDDSNLNPNVNGCDGILNHDVNSLPQVARDLSLOLVNALKGAF
GEVYMALYRHRDGDVEMGVAVKTLREDPKREKEEDFLKEAATMAKFNHNMVHLICV
CFDRQYIYVLELLAGDLOKFLRENNTPERFSLTMTKDLFCALDVAKGCRYMESK
RFTRHDIAARNCLSSKGPGRVVIADFGMSRDYISDYRKRGGKAMLPKRWMPPEAF
LDGIFTSKTDVWSFGILLVEVFSLGRSPYQGHNTQVMELVVRGGRGLSGTECPVSIY
KVMADCNWPTPEDRPTFTLLEHLTACTQDASIMNAPLPNLTPTASERDDTVIRPN
GEEFLAVPDYLVPLPPGSGNPSMAGSGYVPELOROOMSSCTPPAVTSPAAPHPRP
VENIAPTSGDCWETSRLPNSKVEPVVGGSHDVPLAGGEAKLISLDTQPPTTIO
PPLSPASQLDCTILDPSALTSLPNGNAKOSYANQVKNAEAKVINGVNGTVMGNG
DVSTGSGGADSPFTIOGAYERYKDNHNSHISC"
BASE COUNT 2522 a 1856 c 1961 g 2096 t
ORIGIN
Query Match 2.7%; Score 20; DB 5; Length 8435;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 185 tcggcaagaatccggcgagt 204
|||||
Db 2215 TCGCAAGAATCCGGCGAGT 2234
```

```

RESULT 4
AC004287/c
LOCUS
DEFINITION
Drosophila melanogaster DNA sequence (P1 DS02309 (D166)), complete
sequence.
ACCESSION
AC004287 AC003783 AC003310 AC003311 AC003784 AC003312 AC003313
VERSION
AC003314 AC003315 AC003316 AC003317
KEYWORDS
AC004287.1 GI:3769311
SOURCE
Drosophila melanogaster (Subclones in tet from p1 clone DS02309
(D166)) DNA.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 70841)
REFERENCE
1 Celnik, S.E., George, R.A., Galle, R.F., Hoskins, R.A.,
Svirskas, R.R., Harris, N.L., Agbayani, A., Arcaina, T.T., Baxter, E.,
Blazej, R.G., Chavez, C., Chew, M., Doyle, C.M., Farfan, D.E.,
Flanagan, J., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, S.H., Lee, B., Lomotan, M.A., Mak, J., Mazda, P., Moshrefi, A.R.,
Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Punch, E.,
Snir, E., Twomey, B., Wan, K.H., Zhang, R., Zieran, L.L. and Rubin, G.M.
Sequencing of Drosophila chromosome 2R, region 53C10-53D3
Unpublished (1997)
2 (bases 1 to 70841)
AUTHORS
Celnik, S.E., George, R.A., Galle, R., Svirskas, R.R., Hoskins, R.A.,
Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Chavez, C.,
Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A.,
Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B.,
Lomotan, M.A., Mak, J., Mazda, P., Mok, M.S., Moshrefi, A.R.,
Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Punch, E.,
Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R.,
Zieran, L.L. and Kimmel, B.E.
TITLE
Direct Submission
JOURNAL
Submitted (07-MAR-1998) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT
On Oct 19, 1998 this sequence version replaced gi:3766105.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
email to bdg@fruitfly.berkeley.edu.
Library location: 5-25.
FEATURES
Location/Qualifiers
1..70841
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
/map="53C10-53D3"
/clone="P1 DS02309 (D166)"
BASE COUNT 20303 a 15224 c 15085 g 20229 t
ORIGIN
Query Match 2.7%; Score 20; DB 4; Length 70841;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 185 tcggcaagaatccggcagt 204
|||||
DB 48426 TCGGCAAGATCCGGCAGT 48407
|||||

RESULT 5
AC017187
LOCUS
DEFINITION
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
AC017187 93588 bp DNA HTG 09-DEC-1999

```

```

AC017187
AC017187.1 GI:6553799
HTG: HTGS_PHASE2.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 93588)
Adams, M. and Venter, J.C.
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10209809 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..93588
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 26957 a 19517 c 19788 g 27326 t
ORIGIN
Query Match 2.7%; Score 20; DB 64; Length 93588;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 185 tcggcaagaatccggcagt 204
|||||
DB 5700 TCGGCAAGATCCGGCAGT 5719
|||||

RESULT 6
AC003806/c
LOCUS
DEFINITION
Drosophila melanogaster genomic scaffold 142000013386047 section 37
of 52, complete sequence.
ACCESSION
AC003806 AE002787
VERSION
AE003806.2 GI:10727503
KEYWORDS
HTG.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 268219)
Adams, M.D., Celnik, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,
Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D.,
Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor
Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J.,
Andrews-Pfannkuch, C., Baldwin, D., Ballew, R.M., Basu, A.,
Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D.,
Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C.,
Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,
Cherry, J.M., Cavley, S., Dahlke, C., Davenport, L.B., Davies, P., de
Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,
Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,
Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S.,
Fleischmann, W., Flossler, C., Gabrielian, A.E., Garg, N.S.,
Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,
Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,
Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
Kensison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,

```

Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T. C., McLeod, M. P., McPherson, D., Merkulov, G., Milshina, N. V., Mobarly, C., Morris, J., Moshrefi, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L., Nuzny, D. L., Nelson, D. R., Nelson, K. A., Nixon, K., Nusskern, D. R., Pacleb, J. M., Palazzolo, M., Pittman, G. S., Pan, S., Pollard, J., Puri, V., Reese, M. G., Reinert, K., Remington, K., Saunders, R. D., Scheeler, F., Shen, H., Shue, B. C., Siden-Kiamos, I., Simpson, M., Skupski, M. P., Smith, T. I., Spier, E., Spradling, A. C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A. H., Wang, X., Wang, Z. Y., Wassarman, D. A., Weinstein, G. M., Weissbach, J., Williams, S. M., Woodage, T., Worley, K. C., Wu, D., Yang, S., Yao, Q. A., Ye, J., Yeh, R. F., Zaveri, J. S., Zhan, M., Zhang, C., Zhao, Q., Zheng, L., Zheng, X. H., Zhong, F. N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H. O., Gibbs, R. A., Myers, E. W., Rubin, G. M. and Venter, J. C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 268219)
Adams, M. D., Celniker, S. E., Gibbs, R. A., Rubin, G. M. and Venter, C. J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7302890.
Location/Qualifiers
I. .268219
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
<9709. >9957
/gene="CG16797"
/product="CT35933"
/db_xref="FLYBASE:FBan0016797"
/db_xref="FLYBASE:FBgn0034130"
/evidence=not_experimental
<9709. >9957
/gene="CG16797"
/map="53C2-53C3"
/db_xref="FLYBASE:FBan0016797"
/db_xref="FLYBASE:FBgn0034130"
/evidence=not_experimental
9709. 9957
/gene="CG16797"
/note="CG16797 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0016797"
/db_xref="FLYBASE:FBgn0034130"
/evidence=not_experimental
/protein_id="AAF57994.1"
/db_xref="GI:7302922"
/translation="MLELILNVVHLSLQISYEFEIYESISDCSDVPPGTLNISG
LFDLTVTTITADGLSVSGNMTSVFNAQPTDRIEYFA"
complement(join(27841. .27954,28010. .>28540))
/gene="CG15712"
/product="CT35937"
/db_xref="FLYBASE:FBan0015712"
/db_xref="FLYBASE:FBgn0034131"
/evidence=not_experimental
complement(<27841. .>28540)
/gene="CG15712"
/map="53C3-53C4"
/db_xref="FLYBASE:FBan0015712"
/db_xref="FLYBASE:FBgn0034131"
/evidence=not_experimental
complement(join(27841. .27954,28010. .28177))
/gene="CG15712"
/note="CG15712 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0015712"
/db_xref="FLYBASE:FBgn0034131"
/evidence=not_experimental
/protein_id="AAF57993.1"
/db_xref="GI:7302921"

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

mRNA

gene

CDS

mRNA

gene

CDS

mRNA

gene

CDS

mRNA

gene

CDS

mRNA

gene

/translation="MLRRDLNLFWDVHSLPRTRIYSLIAQSKOSTEPFAKLYKALR
DKFKELQVRLSRDLQAPDKKLKSKSINVDEILQIVFEVSWGPKTS"
join(34139. .34418,34498. .36233)
/gene="CG4439"
/product="CT14432"
/db_xref="FLYBASE:FBan0004439"
/db_xref="FLYBASE:FBgn0034132"
<34139. .>36233
/gene="CG4439"
/map="53C3-53C4"
/db_xref="FLYBASE:FBan0004439"
/db_xref="FLYBASE:FBgn0034132"
join(34317. .34418,34498. .36000)
/gene="CG4439"
/note="CG4439 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0004439"
/db_xref="FLYBASE:FBgn0034132"
/protein_id="AAF57992.1"
/db_xref="GI:7302920"
/translation="MTRSQMISLRRSTLGLFRCHKALIFARIPAFQOVCKSEDGSC
NTNLQGVVGLYAKDQKGLKSPGKEFDDRVGKRTITELIKESGLNGELGVRLYQ
NIDKEYWAVAVGLGEGAGFNAEVIDEGMENVVCAVAGARALQOQCTTCHVDGM
EYPOAEGAAMAVRYNVNKKNRITAIPLKLEYSTODDANTRGLFKAESONLARR
LSDTPANQMTSPIPAQATVDALPCGVSVEVRSDMLETQNLNSFLWVAKSGEPPIL
LEVSYCGTSPERPILMLGKGLTFNSGGLCHLPKRGMDIYRGAIVGAACVAAIRAAA
ALSPLINSAVLPCENPSGMATKPGDVVTVLLNGKTMRIKIDISLAGTLDADPLLYA
QYTFKPLVVEGVMASGIRKGCASATGLWTNNSTLWKNFKAGALTGRLWRMPLW
RSTFKLVAPRASYDICTGEGHASSCLAAAILVELVPCSDWVHLDTHCTGMLAKHGVP
PYLLKOCMTGRPTRESIIQFLYQMAK"
complement(join(39980. .42020,42187. .>42430))
/gene="CG7813"
/product="CT23722"
/db_xref="FLYBASE:FBan0007813"
/db_xref="FLYBASE:FBgn0034133"
complement(<39980. .>42430)
/gene="CG7813"
/map="53C3-53C4"
/db_xref="FLYBASE:FBan0007813"
/db_xref="FLYBASE:FBgn0034133"
complement(join(39980. .42020,42187. .42350))
/gene="CG7813"
/note="CG7813 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0007813"
/db_xref="FLYBASE:FBgn0034133"
/protein_id="AAF57991.1"
/db_xref="GI:7302919"
/translation="MWEKLVNCIKSNGNGVKNVSCQVVELADLMKQVPPNQMHKFK
MPAKHEEYKDRVKPYEPMPTIDWEYIRQNVREEFVDWVKGYETKYDLHSVFENRH
AIVDHRKYFLVDEEKVVTKCISEYKAESDKRIQELTEKLEFVKAMRPSYEMTMEEF
CFARHLPADPFINKPTFWPTPEQMPGSDPDAALAHHEEPPEPKPKPEKLPDK
PSGEGKPPAAAPKEPKPEVVDTSQAEKATLIAKDLMAKAIVLNLSLKKKMSGLAK
VOKKADAAKAARSETASKSATPAATKSLDSITERESGNICNOTIIRSEAEANPE
VKARHTNLSLEADPCDQAEAREARTLRRKQKQKAEWEKHKTKKDCDKEFTD
PCKLMEETVCKPDKCAKEEAVCPDQCKEEDDICKPDKPCKPDEGSGEDDE
DPCKKPKCDGYQFGDNEGGSQKQDQVFINISESKELAKPEKDESGSGRKP
EKASLSIPQGSQKPVLMQITESKSKDKHKAIEGANEIGPVYTDPNQALVELYQ
KKEKPVIIIEAOKPEDDKPITIVPKLEISANPKGDQVQKQAEQKSPKDMAKQVFTWA
SGAATLLTEATNTLEDLKRKEARLEALEQATYSAQKQAGALAEASKAYEAAANLAI
RSAEQTGEVSRRDLEMAEKHAILAKAGRAVALKDEIARVLDKLKKK"
join(59206. .59270,60019. .60341,68374. .68461,68713. .68764,
70614. .70682,71542. .71649,71899. .72036,72108. .72240,
72359. .72715,73250. .73478,73729. .74065,74232. .74395,
74459. .74611,74794. .74902,75329. .75664)
/gene="Sema-2a"
/note="Nucleotide sequence of the Cclera sequence differs
from the published sequence of this transcript."
/product="CT42076"
/db_xref="FLYBASE:FBan0004700"
/db_xref="FLYBASE:FBgn0011260"
<59206. .>75664
/gene="Sema-2a"

```

/note="CG4700"
/map="53C4-53C6"
/db_xref="FLYBASE:FBan0004700"
join(59220..59270,59920..60341,68374..68461,68713..68764,
70614..70682,71542..71649,71899..72036,72108..72240,
72359..72715,73250..73478,73729..74065,74232..74395,
74459..74611,74794..74902,75329..75664)
/gene="Sema-2a"
/product="CT14922"
/db_xref="FLYBASE:FBan0004700"
join(60176..60341,68374..68461,68713..68764,70614..70682,
71542..71649,71899..72036,72108..72240,72359..72715,
73250..73478,73729..74065,74232..74395,74459..74611,
74794..74902,75329..75400)
/gene="Sema-2a"
/note="Sema-2a gene product [alt 1]"
/codon_start=1

Query Match      2.7%; Score 20; DB 5; Length 268219;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 tcggcaagaatccggcagc 204
|||||
Db 175024 TCGGCAAGATCGGGCAGT 175005

RESULT 7
LOCUS HSU45983 1944 bp DNA PRI 01-JUL-1997
DEFINITION Homo sapiens CCR8 chemokine receptor (CKBR8) gene, complete cds.
ACCESSION U45983.1 GI:2231165
VERSION U45983.1
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1944)
AUTHORS Tiffany,H.L., Lautens,L.L., Gao,J.-L., Pease,J., Locati,M.,
Combadere,C., Modi,W., Bonner,T.I. and Murphy,P.M.
TITLE Identification of a human monocyte and thymus receptor for the CC
chemokine I-309
JOURNAL J. Exp. Med. (1997) In press
REFERENCE 2 (bases 1 to 1944)
AUTHORS Bonner,T.I.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1996) Tom I. Bonner, Lab of Cell Biology, NIMH,
Bldg 36, Rm 3A-17, MSC 4090, Bethesda, MD 20892-4090, USA
REFERENCE 3 (bases 1 to 1944)
AUTHORS Bonner,T.I.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1997) Tom I. Bonner, Lab of Cell Biology, NIMH,
Bldg 36, Rm 3A-17, MSC 4090, Bethesda, MD 20892-4090, USA
REMARK Sequence update by submitter
COMMENT On Jul 1, 1997 this sequence version replaced gi:1245056.
FEATURES
source
1..1944
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p22-p22; between D3S1260 and D3S322"
<1..243
/note="based on comparison to GenBank Accession Number
AF005210"
evidence=not_experimental
gene 258..1325
/gene="CKBR8"
CDS 258..1325
/gene="CKBR8"
/note="GPR-CY6; G protein-coupled receptor"

/codon_start=1
/product="CCR8 chemokine receptor"
/protein_id="AAB61962.1"
/db_xref="GI:1245057"
/translation="MDYTLDSLVTVTVDYYPDISSPCDAELIQTNGKLLAVFYCL
LFVFSLLGNSLIVLVVCKKLSITDVLNLALSDLLFVSPFQTYLLDQWVFG
TMVKVSGFYIGFYSSMFFITLMSVDRLAVHVAVALKVRTIRMGTTLCIAVLT
AIMATIPLLVYQVASEDGLQCYSFYNQOTLKKWIFTNFKNIGLLIPIPTIFECY
IKILHLKRCQNHKTKAIRLVLIIVIASLLFWPFPNVVLFTLSLHSHILDGCSISQ
QLTYATHVTEIISFTHCCVNPVIYAFVGEKFKHLSLSEIFQKSCSIFNYLGRMPRES
CEKSSSCQHSRSSSSVDYL"
547 a 422 c 410 g 565 t

BASE COUNT
ORIGIN
Query Match      2.6%; Score 19; DB 97; Length 1944;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 agaagtgatattgaatt 183
|||||
Db 1644 AGAAGTGGATATTGAAGTT 1662

RESULT 8
LOCUS HSU62556 2608 bp DNA PRI 02-AUG-1996
DEFINITION Human chemokine receptor-like protein (TER1) gene, complete cds.
ACCESSION U62556
VERSION U62556.1 GI:1468978
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 2608)
AUTHORS Napolitano,M., Zingoni,A., Bernardini,G., Spinetti,G., Rocchi,M.
and Santoni,A.
TITLE Human chemokine receptor-like TER1 gene
JOURNAL J. Immunol. (1996) In press
REFERENCE 2 (bases 1 to 2608)
AUTHORS Napolitano,M., Zingoni,A., Bernardini,G., Spinetti,G., Rocchi,M.
and Santoni,A.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-1996) Laboratory of Pathophysiology, Regina Elena
Cancer Institute, via delle Messi d'Oro 156, Rome 00158, Italy
FEATURES
source
1..2608
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p21"
/rpt_family="Alu"
repeat_region 1..142
gene 361..2200
/gene="TER1"
CDS 361..1428
/gene="TER1"
/note="Seven trans membrane spanning chemokine-like
receptor"
/codon_start=1
/product="chemokine receptor-like protein"
/protein_id="AAB05542.1"
/db_xref="GI:1468979"

intron
1..1944
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p22-p22; between D3S1260 and D3S322"
<1..243
/note="based on comparison to GenBank Accession Number
AF005210"
evidence=not_experimental
gene 258..1325
/gene="CKBR8"
CDS 258..1325
/gene="CKBR8"
/note="GPR-CY6; G protein-coupled receptor"

```

```

polyA_signal      2149..2155
                  /note="putative"
                  /gene="TER1"
polyA_signal      2156..2161
                  /note="putative"
                  /gene="TER1"
polyA_signal      2195..2200
                  /note="putative"
                  /gene="TER1"
BASE COUNT      750 a 552 c 523 g 783 t
ORIGIN

```

Query Match 2.6%; Score 19; DB 97; Length 2608;

Best Local Similarity 100.0%; Pred. No. 95;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 agaagtggtattgaagtt 183

|||||

Db 1744 AGAAGTGGATATTGAAGTT 1762

RESULT 9

HSU62556/c

LOCUS HSU62556 2608 bp DNA PRI 02-AUG-1996

DEFINITION Human chemokine receptor-like protein (TER1) gene, complete cds.

ACCESSION U62556

VERSION U62556.1 GI:1468978

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2608)

AUTHORS Napolitano,M., Zingoni,A., Bernardini,G., Spinetti,G., Rocchi,M.

and Santoni,A.

TITLE Human chemokine receptor-like TER1 gene

JOURNAL Immunol. (1996) In press

REFERENCE 2 (bases 1 to 2608)

AUTHORS Napolitano,M., Zingoni,A., Bernardini,G., Spinetti,G., Rocchi,M.

and Santoni,A.

TITLE Direct Submission

JOURNAL Submitted (28-JUN-1996) Laboratory of Pathophysiology, Regina Elena

Cancer Institute, via delle Messi d'Oro 156, Rome 00158, Italy

FEATURES

Source

1..2608

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="3"

/map="3p21"

/repeat_region

1..142

/rpt_family="Alu"

361..2200

/gene="TER1"

361..1428

/gene="TER1"

/note="Seven trans-membrane spanning chemokine-like

receptor"

/codon_start=1

/product="chemokine receptor-like protein"

/protein_id="AAB05542.1"

/db_xref="GI:1468978"

/translation="MDYTLDSLVTVTDYYYPDIFSSPCDAELIOTNGKLLAVFVCL

LFVSLGNSLVILVWCKKLRSITDYLLNALSDLFLVFSFFPTQYLLDQWFG

TVWCVSGFYIGFTSSMFFITLMSVDRLAVHVAALKVRTIRMGTTCLAVWLT

AIMATPLLVFVAVASEDGVLCQYSFYNQOTLKWIFNFRNMLGLLPIFTMFCY

IKLHQLKRCNNHKTAKRLVLIIVIASLLEWPFNVVFLTSLHSHILDGCSISQ

QUTYATHTVEIISFTHCVCNPIYAVFGVKFKKHLSEIFQKSCSQIFNVYLGMPRES

CEKSSSCQHQSSRSSVDYIL"

2143..2148

/gene="TER1"

/note="putative"

polyA_signal

```

polyA_signal      2149..2155
                  /gene="TER1"
                  /note="putative"
polyA_signal      2156..2161
                  /gene="TER1"
                  /note="putative"
polyA_signal      2195..2200
                  /gene="TER1"
                  /note="putative"
BASE COUNT      750 a 552 c 523 g 783 t
ORIGIN

```

Query Match 2.6%; Score 19; DB 97; Length 2608;

Best Local Similarity 100.0%; Pred. No. 95;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 agaagtggtattgaagtt 183

|||||

Db 2544 AGAAGTGGATATTGAAGTT 2526

RESULT 10

AY016370

LOCUS AY016370 86172 bp DNA PRI 21-FEB-2001

DEFINITION Homo sapiens hmrNPAL pseudogene, complete sequence; and CC

chemokine receptor 8 (CCR8) and CX3C chemokine receptor 1 (CX3CR1)

genes, complete cds.

ACCESSION AY016370

VERSION AY016370.1 GI:13027668

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 86172)

AUTHORS DeVries,M.E., Xu,L., Kelvin,A.A. and Kelvin,D.J.

TITLE Identification and characterization of the CX3CR1 and CCR8

promoters

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 86172)

AUTHORS Xu,L., DeVries,M.E., Kelvin,A.A. and Kelvin,D.J.

TITLE Direct Submission

JOURNAL Submitted (05-DEC-2000) Laboratory of Molecular Inflammation and

Immunology, Robarts Research Institute, 1400 Western Rd, London, ON

N6G 2V4, Canada

FEATURES Location/Qualifiers

Source

1..86172

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="3"

14331..15366

/gene="hmrNPAL"

/pseudo

complement(join(16138..17996,20486..20608))

/gene="CCR8"

/note="transcript 1"

/product="CC chemokine receptor 8"

complement(16138..18030)

/gene="CCR8"

/note="transcript 2"

/product="CC chemokine receptor 8"

complement(16138..20608)

/gene="CCR8"

complement(16138..16143)

/gene="CCR8"

/note="putative"

complement(16177..16182)

/gene="CCR8"

/note="putative"

complement(16183..16189)

/gene="CCR8"

/note="putative"

polyA_signal

```
polyA_signal complement(16190. .16195)
/gene="CCR8"
/note="putative"
CDS complement(16915. .17982)
/gene="CCR8"
/codon_start=1
/product="CC chemokine receptor 8"
/protein_id="AAK08628.1"
/db_xref="GI:13027670"
/translation="MDYTLDSLVTVTDDYYPDIFFSPDCGELIQTNGKLLAVFYCL
LFVSLGNSLVILVVCCKLRSITDYLNLALSDLLFVFPFQTYILLDQWVFG
TMVKVSGFYIGFYSMEFTILMSVDYLAHVAVALKVIRTIMGTTICLAWLT
AIMATIPLLVFOVASSEDVLCQCYSFYNQOQTLKWKIFTNFKMNLGLLIPFTIMFCY
IKILHQLKRCQNHKTKAIRLVIVIASLLFVFPNVVFLTSLHSHHLDGCSISQ
OVLVATHVTEIISFTHCCNVPVIYAFVGEKPKHLSLSEIFQKSCQIFNVLGRQMPRES
CEKSSSCQOHSRSSVDYL"
join(68700. .68784,83882. .>86172)
/gene="CX3CR1"
/note="transcript 1"
/product="CX3C chemokine receptor 1"
/evidence=experimental
68700. .>86172
/gene="CX3CR1"
join(70361. .70441,83882. .>86172)
/note="transcript 2"
/product="CX3C chemokine receptor 1"
/evidence=experimental
83891. .84958
/gene="CX3CR1"
/codon_start=1
/product="CX3C chemokine receptor 1"
/protein_id="AAK08627.1"
/db_xref="GI:13027669"
/translation="MDQFPESTVENFEVDLAECYIGDIIVFGTVFLSIFYSVIFAI
GLVGNLVFALTSKPKSVTDIYLLNLALSDLLFVATLPFWTHYLINKEGLHNCM
KFTTAFPIFGSIFITVISIDRYLAIVLAANSNNRTVQHGVTISLGYAAAILV
AAPQFMFKQKNECLGDYPEVLOEIWPLRNVTNPLGLLPLIMSYCFRIQTL
FSCNHHKAKIKILIVIVFLEFWPYNMIETLKLVDFFPSCDMKDLRLALS
VTEVAFSHCCNPLIYAFAGEKFRYLYHLGYKCLAVLCGRSVHVDFFSSSESORSRH
GSVLSNFTVHTSDGALLL"
BASE COUNT 23130 a 18485 c 17978 g 26342 t 237 others
ORIGIN
Query Match 2.6%; Score 19; DB 91; Length 86172;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 165 agaagtggatattgaagtt 183
|||||
Db 15794 AGAAGTGGATATTGAAGTT 15812

RESULT 11
AY016370/c
LOCUS AY016370.1 86172 bp DNA PRI 21-FEB-2001
DEFINITION Homo sapiens hnrnpal pseudogene, complete sequence; and CC
chemokine receptor 8 (CCR8) and CX3C chemokine receptor 1 (CX3CR1)
genes, complete cds.
ACCESSION AY016370
VERSION AY016370.1 GI:13027668
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 86172)
AUTHORS DeVries, M.E., Xu, L., Kelvin, A.A. and Kelvin, D.J.
TITLE Identification and characterization of the CX3CR1 and CCR8
promoters
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 86172)
```

Xu, L., DeVries, M.E., Kelvin, A.A. and Kelvin, D.J.
Direct Submission
Submitted (05-DEC-2000) Laboratory of Molecular Inflammation and
Immunology, Robarts Research Institute, 1400 Western Rd, London, ON
N6G 2V4, Canada
Location/Qualifiers
1. .86172
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
14331. .15366
/gene="hnrnpal"
/pseudo
complement(join(16138. .17996,20486. .20608))
/gene="CCR8"
/note="transcript 1"
/product="CC chemokine receptor 8"
complement(16138. .18030)
/gene="CCR8"
/note="transcript 2"
/product="CC chemokine receptor 8"
complement(16138. .20608)
/gene="CCR8"
complement(16138. .16143)
/gene="CCR8"
/note="putative"
complement(16177. .16182)
/gene="CCR8"
/note="putative"
complement(16183. .16189)
/gene="CCR8"
/note="putative"
complement(16190. .16195)
/gene="CCR8"
/note="putative"
complement(16915. .17982)
/gene="CCR8"
/codon_start=1
/product="CC chemokine receptor 8"
/protein_id="AAK08628.1"
/db_xref="GI:13027670"
/translation="MDYTLDSLVTVTDDYYPDIFFSPDCGELIQTNGKLLAVFYCL
LFVSLGNSLVILVVCCKLRSITDYLNLALSDLLFVFPFQTYILLDQWVFG
TMVKVSGFYIGFYSMEFTILMSVDYLAHVAVALKVIRTIMGTTICLAWLT
AIMATIPLLVFOVASSEDVLCQCYSFYNQOQTLKWKIFTNFKMNLGLLIPFTIMFCY
IKILHQLKRCQNHKTKAIRLVIVIASLLFVFPNVVFLTSLHSHHLDGCSISQ
OVLVATHVTEIISFTHCCNVPVIYAFVGEKPKHLSLSEIFQKSCQIFNVLGRQMPRES
CEKSSSCQOHSRSSVDYL"
join(68700. .68784,83882. .>86172)
/gene="CX3CR1"
/note="transcript 1"
/product="CX3C chemokine receptor 1"
/evidence=experimental
68700. .>86172
/gene="CX3CR1"
join(70361. .70441,83882. .>86172)
/gene="CX3CR1"
/note="transcript 2"
/product="CX3C chemokine receptor 1"
/evidence=experimental
83891. .84958
/gene="CX3CR1"
/codon_start=1
/product="CX3C chemokine receptor 1"
/protein_id="AAK08627.1"
/db_xref="GI:13027669"
/translation="MDQFPESTVENFEVDLAECYIGDIIVFGTVFLSIFYSVIFAI
GLVGNLVFALTSKPKSVTDIYLLNLALSDLLFVATLPFWTHYLINKEGLHNCM
KFTTAFPIFGSIFITVISIDRYLAIVLAANSNNRTVQHGVTISLGYAAAILV
AAPQFMFKQKNECLGDYPEVLOEIWPLRNVTNPLGLLPLIMSYCFRIQTL
FSCNHHKAKIKILIVIVFLEFWPYNMIETLKLVDFFPSCDMKDLRLALS
VTEVAFSHCCNPLIYAFAGEKFRYLYHLGYKCLAVLCGRSVHVDFFSSSESORSRH
GSVLSNFTVHTSDGALLL"

BASE COUNT 23130 a 18485 c 17978 g 26342 t 237 others.
ORIGIN

Query Match 2.6%; Score 19; DB 91; Length 86172;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 165 agaagtgtattgaagt 183
|||||

Db 16594 AGAAGTGGATATTGAAGTT 16576

RESULT 12

AL136961/c

LOCUS

DEFINITION

AL136961

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

requests: clones@esanger.ac.uk

On Apr 12, 2001 this sequence version replaced gi:13276997.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em., EMBL; Sw.,

SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP

database can be found at

<http://www.sanger.ac.uk/Projects/C-elegans/wormpep> This sequence

was generated from part of bacterial clone contigs of human

chromosome 13, constructed by the Sanger Centre Chromosome 13

Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr13>

RP11-178C10 is from the library RPC1-11.1 constructed by the group

of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone

RP11-178C10. It may be shorter because we sequence overlapping

sections only once, except for a 100 base overlap.

The true right end of clone RP11-178C10 is at 98861 in this

sequence. The true right end of clone RP11-461N23 is at 100 in this

sequence.

FEATURES

Location/Qualifiers

1..98861

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="13"

/clone="RP11-178C10"

/clone.lib="RPC1-11.1"

161..463

/note="LIMB8 repeat: matches 5836..6160 of consensus"

repeat_region

161..463

/note="LIMB8 repeat: matches 5836..6160 of consensus"

repeat_region

2385..2420

/note="L2 repeat: matches 2592..2659 of consensus"

repeat_region

4555..4646

/note="MIR repeat: matches 111..146 of consensus"

repeat_region

4989..5306

/note="L2 repeat: matches 1031..1120 of consensus"

repeat_region

5859..5987

/note="MER7A repeat: matches 2..345 of consensus"

repeat_region

6215..6548

/note="L2 repeat: matches 2574..2708 of consensus"

repeat_region

6557..6635

/note="Charliel repeat: matches 2292..2629 of consensus"

repeat_region

6730..7233

/note="L2 repeat: matches 2128..2206 of consensus"

repeat_region

8301..8499

/note="L2 repeat: matches 1330..1889 of consensus"

repeat_region

8893..8994

/note="MIR repeat: matches 17..219 of consensus"

repeat_region

9079..9126

/note="L2 repeat: matches 1551..1656 of consensus"

repeat_region

9554..9754

/note="24 copies 2 mer at 75% conserved"

repeat_region

9816..9979

/note="MER20 repeat: matches 33..218 of consensus"

repeat_region

10402..10703

/note="FAM repeat: matches 2..165 of consensus"

repeat_region

12181..12626

/note="LIM1 repeat: matches 5364..5577 of consensus"

repeat_region

12654..13012

/note="L1 repeat: matches 3474..3949 of consensus"

repeat_region

13491..14076

/note="LIME3 repeat: matches 5571..5927 of consensus"

repeat_region

16310..17185

/note="L1 repeat: matches 2226..2881 of consensus"

repeat_region

17207..17622

/note="LIMC4 repeat: matches 6237..7151 of consensus"

repeat_region

17628..18380

/note="MER47B repeat: matches 2..413 of consensus"

repeat_region

20797..21100

/note="LIMC4 repeat: matches 7197..7976 of consensus"

repeat_region

22146..22814

/note="ALU repeat: matches 1..297 of consensus"

repeat_region

23758..23981

/note="L2 repeat: matches 2574..2643 of consensus"

repeat_region

25108..25212

/note="ALU repeat: matches 85..310 of consensus"

repeat_region

25740..25795

/note="L2 repeat: matches 2570..2688 of consensus"

repeat_region

25982..26012

/note="MIR repeat: matches 196..251 of consensus"

26005..26208
/note="Tigger4(Zombi) repeat: matches 2701..2731 of consensus"
26209..26513
/note="Alus repeat: matches 1..305 of consensus"
26514..28201
/note="Tigger4(Zombi) repeat: matches 1..1772 of consensus"
32616..32843
/note="6 copies 38 mer 86% conserved"
35333..35362
/note="15 copies 2 mer ac 100% conserved"
36220..36303
/note="2 copies 42 mer 97% conserved"
36967..37274
/note="Alus repeat: matches 1..313 of consensus"
39481..39767
/note="Alus repeat: matches 1..297 of consensus"
41084..41111
/note="14 copies 2 mer aa 89% conserved"

```

repeat_region 41119..41174
/note="THE1C repeat: matches 1. .58 of consensus"
repeat_region 41175..41482
/note="AluSp repeat: matches 1. .309 of consensus"
repeat_region 41483..41785
/note="THE1C repeat: matches 58. .344 of consensus"
repeat_region 41791..41842
/note="L2 repeat: matches 1215. .1266 of consensus"
repeat_region 42051..42573
/note="MER74A repeat: matches 2. .555 of consensus"
repeat_region 42857..43014
/note="L2 repeat: matches 2553. .2733 of consensus"
repeat_region 43250..43279
/note="L5 copies 2 mer ca 100% conserved"
repeat_region 44262..44708
/note="MLT1C repeat: matches 1. .466 of consensus"
repeat_region 44860..45175
/note="MLT2FB repeat: matches 1. .328 of consensus"
repeat_region 45204..45235
/note="L6 copies 2 mer tg 87% conserved"
repeat_region 45389..45559
/note="MER5B repeat: matches 4. .175 of consensus"
repeat_region 45639..45788
/note="MER5A repeat: matches 9. .184 of consensus"
repeat_region 45963..46162
/note="MIR repeat: matches 25. .250 of consensus"
repeat_region 46281..46837
/note="MER67D repeat: matches 1. .510 of consensus"
repeat_region 47886..51275
/note="L1PA6 repeat: matches 2744. .6141 of consensus"
repeat_region 51364..51523
/note="MER5A repeat: matches 1. .189 of consensus"
repeat_region 51736..51791
/note="MIR repeat: matches 118. .178 of consensus"
repeat_region 51848..52167
/note="MSTB repeat: matches 86. .422 of consensus"
repeat_region 53138..53915
/note="MER21B repeat: matches 41. .789 of consensus"
repeat_region 54321..54629
/note="AluSg repeat: matches 1. .310 of consensus"
repeat_region 55247..55308
/note="MER69A repeat: matches 24. .87 of consensus"
repeat_region 55309..55622
/note="AluSg repeat: matches 1. .313 of consensus"
repeat_region 55623..55715
/note="MER5A repeat: matches 87. .176 of consensus"
repeat_region 57171..57271
/note="MLT1F repeat: matches 437. .541 of consensus"
repeat_region 57462..57560
/note="MLT1G repeat: matches 169. .264 of consensus"
repeat_region 57561..58049
/note="MER74A repeat: matches 1. .519 of consensus"
repeat_region 58063..58287
/note="AluSg/x repeat: matches 163. .300 of consensus"
repeat_region 58336..58475
/note="MLT1G repeat: matches 76. .205 of consensus"
repeat_region 58496..58809
/note="AluSx repeat: matches 1. .311 of consensus"
repeat_region 58818..59182
/note="L1ME3A repeat: matches 5789. .6157 of consensus"
repeat_region 59198..59336
/note="MER74A repeat: matches 1. .139 of consensus"
repeat_region 59337..59435
/note="MLT1G repeat: matches 169. .264 of consensus"
repeat_region 59626..59726
/note="MLT1F repeat: matches 437. .541 of consensus"
repeat_region 59804..60249
/note="L1ME3A repeat: matches 5301. .5782 of consensus"
repeat_region 61972..62009

```

Query Match 2.6%, Score 19; DB 89; Length 98861;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 165 agaagtggatattgaatt 183
 |||||
 Db 26530 AGAAGTGGATATTGAAGTT 26512

RESULT 13
 AC084179
 LOCUS
 DEFINITION
 AC084179
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AC084179 147009 bp DNA HTG 15-JAN-2001
 Homo sapiens chromosome 5 clone RP11-148P16 map 5, WORKING DRAFT
 SEQUENCE, 39 unordered pieces.
 AC084179
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 human

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

1 (bases 1 to 147009)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 5, clone RP11-148P16
 Unpublished
 2 (bases 1 to 147009)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bida,F., Boguslavsky,L.,
 Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 DeRellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
 FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
 Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
 Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,
 Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
 McSheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
 Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
 O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,
 Pierre,N., Pisan,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
 Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
 Sougnaz,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Tirrell,A., Travers,M., Triggilio,J., Vassiliev,H., Viel,R., Vo.A.,
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
 Zimmer,A. and Zody,M.

TITLE
 JOURNAL
 COMMENT

Direct Submission
 Submitted (14-OCT-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jan 15, 2001 this sequence version replaced gi:10801407.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 Project Information
 Center project name: L11428
 Center clone name: 148_P_16

Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 13223 bases at least Q40
 Consensus quality: 139576 bases at least Q30
 Consensus quality: 142136 bases at least Q20
 Insert size: 176000; agarose-fp
 Insert size: 143209; sum-of-contigs
 Quality coverage: 3.0 in Q20 bases; agarose-fp
 Quality coverage: 3.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 39 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1353: contig of 1353 bp in length
1354 1453: gap of 100 bp
1454 2001: contig of 548 bp in length
2002 2101: gap of 100 bp
2102 3124: contig of 1023 bp in length
3125 3224: gap of 100 bp
3225 4546: contig of 1322 bp in length
4547 4646: gap of 100 bp
4647 6015: contig of 1369 bp in length
6016 6115: gap of 100 bp
6116 7299: contig of 1184 bp in length
7300 7399: gap of 100 bp
7400 8836: contig of 1437 bp in length
8837 8936: gap of 100 bp
8937 10282: contig of 1346 bp in length
10283 10382: gap of 100 bp
10383 11352: contig of 970 bp in length
11353 11452: gap of 100 bp
11453 12856: contig of 1404 bp in length
12857 12956: gap of 100 bp
12957 14187: contig of 1231 bp in length
14188 14287: gap of 100 bp
14288 15977: contig of 1690 bp in length
15978 16077: gap of 100 bp
16078 17844: contig of 1767 bp in length
17845 17944: gap of 100 bp
17945 19861: contig of 1917 bp in length
19862 19961: gap of 100 bp
19962 22368: contig of 2407 bp in length
22369 22468: gap of 100 bp
22469 24485: contig of 2017 bp in length
24486 24585: gap of 100 bp
24586 26722: contig of 2137 bp in length
26723 26822: gap of 100 bp
26823 28983: contig of 2161 bp in length
28984 29083: gap of 100 bp
29084 31858: contig of 2775 bp in length
31859 31958: gap of 100 bp
31959 33783: contig of 1825 bp in length
33784 33883: gap of 100 bp
33884 36259: contig of 2376 bp in length
36260 36359: gap of 100 bp
36360 39396: contig of 3037 bp in length
39397 39496: gap of 100 bp
39497 43329: contig of 3833 bp in length
43330 43429: gap of 100 bp
43430 46057: contig of 2628 bp in length
46058 46157: gap of 100 bp
46158 50066: contig of 3909 bp in length
50067 50166: gap of 100 bp
50167 53727: contig of 3561 bp in length
53728 53827: gap of 100 bp
53828 57631: contig of 3804 bp in length
57632 57731: gap of 100 bp
57732 61304: contig of 3573 bp in length
61305 61404: gap of 100 bp
61405 63868: contig of 2464 bp in length
63869 63968: gap of 100 bp
63969 68393: contig of 4425 bp in length
68394 68493: gap of 100 bp
68494 72466: contig of 3973 bp in length
72467 72566: gap of 100 bp
72567 77027: contig of 4461 bp in length
77028 77127: gap of 100 bp
77128 81524: contig of 4397 bp in length
81525 81624: gap of 100 bp
81625 88873: contig of 7249 bp in length
88874 88973: gap of 100 bp

* 88974 96287: contig of 7314 bp in length
96288 96387: gap of 100 bp
96388 117899: contig of 21512 bp in length
117900 117999: gap of 100 bp
118000 123663: contig of 5664 bp in length
123664 123763: gap of 100 bp
123764 137598: contig of 13835 bp in length
137599 137698: gap of 100 bp
137699 147009: contig of 9311 bp in length.

FEATURES

source

1. 147009
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5"
/clone="RP11-148P16"
/clone_lib="RPC1-11 Human Male BAC"
1. 1353
/note="assembly_fragment"
clone_end:SP6
vector_side:left
1454..2001
/note="assembly_fragment"
2102..3124
/note="assembly_fragment"
3225..4546
/note="assembly_fragment"
4647..6015
/note="assembly_fragment"
6116..7299
/note="assembly_fragment"
7400..8836
/note="assembly_fragment"
8937..10282
/note="assembly_fragment"
10383..11352
/note="assembly_fragment"
11453..12856
/note="assembly_fragment"
12957..14187
/note="assembly_fragment"
14288..15977
/note="assembly_fragment"
16078..17844
/note="assembly_fragment"
17945..19861
/note="assembly_fragment"
19962..22368
/note="assembly_fragment"
22469..24485
/note="assembly_fragment"
24586..26722
/note="assembly_fragment"
26823..28983
/note="assembly_fragment"
29084..31858
/note="assembly_fragment"
31959..33783
/note="assembly_fragment"
33884..36259
/note="assembly_fragment"
36360..39396

Query Match 2.6% Score 19; DB 76; Length 147009;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 111 ttccatgttccctaccag 129

|||||

Db 79601 TTCCATGTTCCCTACCAG 79619

RESULT 14

AC016189

LOCUS AC016189 151174 bp DNA HTG 13-JUL-2000
 DEFINITION Homo sapiens clone RP11-23M9, LOW-PASS SEQUENCE SAMPLING.
 AC016189
 AC016189 3 GI:9129134
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1. (bases 1 to 151174)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

JOURNAL

Unpublished

REFERENCE

2. (bases 1 to 151174)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouckghalter,B.,

Brown,A., Castle,A., Collangelo,M., Collins,S., Collymore,A.,

Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,

Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,

Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,

McEwan,P., McCurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,

Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,

Wyman,B., Ye,W.J., Zimmer,A. and Zody,M.

TITLE

JOURNAL

Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:6939915.

COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4399

Center clone name: 23_M_9

* NOTE: This record contains 152 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

* 1

* 852 951: contig of 851 bp in length

* 952 1804: contig of 853 bp in length

* 1805 1904: gap of 100 bp

* 1905 2719: contig of 815 bp in length

* 2720 2819: gap of 100 bp

* 2820 3668: contig of 849 bp in length

* 3669 3768: gap of 100 bp

* 3769 4657: contig of 889 bp in length

* 4658 4757: gap of 100 bp

* 4758 5600: contig of 843 bp in length

* 5601 5700: gap of 100 bp

* 5701 6532: contig of 832 bp in length

* 6533 6632: gap of 100 bp

* 6633 7481: contig of 849 bp in length

* 7482 7581: gap of 100 bp

* 7582 8447: contig of 866 bp in length

* 8448 8547: gap of 100 bp

* 8548 9383: contig of 836 bp in length

* 9384 9483: gap of 100 bp

* 9484 10342: contig of 859 bp in length

* 10343 10442: gap of 100 bp

* 10443 11223: contig of 881 bp in length

* 11224 11423: gap of 100 bp

* 11424 12254: contig of 831 bp in length

* 12255 12354: gap of 100 bp

* 12355 13288: contig of 934 bp in length

* 13289 13388: gap of 100 bp

* 13389 14235: contig of 847 bp in length

* 14236 14335: gap of 100 bp

* 14336 15172: contig of 837 bp in length

* 15173 15272: gap of 100 bp

* 15273 16130: contig of 858 bp in length

* 16131 16230: gap of 100 bp

* 16231 17073: contig of 843 bp in length

* 17074 17173: gap of 100 bp

* 17174 18003: contig of 830 bp in length

* 18004 18103: gap of 100 bp

* 18104 19339: contig of 836 bp in length

* 18940 19039: gap of 100 bp

* 19040 19885: contig of 846 bp in length

* 19886 19985: gap of 100 bp

* 19986 20834: contig of 849 bp in length

* 20835 20934: gap of 100 bp

* 20935 21761: contig of 827 bp in length

* 21762 21861: gap of 100 bp

* 21862 22743: contig of 882 bp in length

* 22744 22843: gap of 100 bp

* 22844 23708: contig of 865 bp in length

* 23709 23808: gap of 100 bp

* 23809 24676: contig of 868 bp in length

* 24677 24776: gap of 100 bp

* 24777 25602: contig of 826 bp in length

* 25603 25702: gap of 100 bp

* 25703 26560: contig of 858 bp in length

* 26561 26660: gap of 100 bp

* 26661 27504: contig of 844 bp in length

* 27505 27604: gap of 100 bp

* 27605 28449: contig of 845 bp in length

* 28450 28549: gap of 100 bp

* 28550 29398: contig of 849 bp in length

* 29399 29498: gap of 100 bp

* 29499 30357: contig of 859 bp in length

* 30358 30457: gap of 100 bp

* 30458 31323: contig of 866 bp in length

* 31324 31423: gap of 100 bp

* 31424 32300: contig of 877 bp in length

* 32301 32400: gap of 100 bp

* 32401 33254: contig of 854 bp in length

* 33255 33354: gap of 100 bp

* 33355 34218: contig of 864 bp in length

* 34219 34318: gap of 100 bp

* 34319 35169: contig of 851 bp in length

* 35170 35269: gap of 100 bp

* 35270 36125: contig of 856 bp in length

* 36126 36225: gap of 100 bp

* 36226 37061: contig of 836 bp in length

* 37062 37161: gap of 100 bp

* 37162 38003: contig of 842 bp in length

* 38004 38103: gap of 100 bp

* 38104 38945: contig of 842 bp in length

* 38946 39045: gap of 100 bp

* 39046 39882: contig of 837 bp in length

* 39883 39982: gap of 100 bp

* 39983 40833: contig of 851 bp in length

* 40834 40933: gap of 100 bp

* 40934 41773: contig of 846 bp in length

* 41780 41879: gap of 100 bp

* 41880 42722: contig of 843 bp in length

* 42723 42822: gap of 100 bp

```

* 42823 43678: contig of 856 bp in length
* 43679 43778: gap of 100 bp
* 43779 44634: contig of 856 bp in length
* 44635 44734: gap of 100 bp
* 44735 45587: contig of 853 bp in length
* 45588 45687: gap of 100 bp
* 45688 46542: contig of 855 bp in length
* 46543 46642: gap of 100 bp
* 46643 47488: contig of 846 bp in length
* 47489 47588: gap of 100 bp
* 47589 48448: contig of 860 bp in length
* 48449 48548: gap of 100 bp
* 48549 49450: contig of 902 bp in length
* 49451 49550: gap of 100 bp
* 49551 50383: contig of 833 bp in length
* 50384 50483: gap of 100 bp
* 50484 51353: contig of 870 bp in length
* 51354 51453: gap of 100 bp
* 51454 52308: contig of 853 bp in length
* 52309 52408: gap of 100 bp
* 52409 53252: contig of 846 bp in length
* 53253 53352: gap of 100 bp
* 53353 54224: contig of 872 bp in length
* 54225 54324: gap of 100 bp
* 54325 55162: contig of 838 bp in length
* 55163 55262: gap of 100 bp
* 55263 56086: contig of 824 bp in length
* 56087 56186: gap of 100 bp
* 56187 57026: contig of 840 bp in length
* 57027 57126: gap of 100 bp
* 57127 58025: contig of 899 bp in length
* 58026 58125: gap of 100 bp
* 58126 59004: contig of 879 bp in length
* 59005 59104: gap of 100 bp
* 59105 59979: contig of 875 bp in length
* 59980 60079: gap of 100 bp
* 60080 60920: contig of 841 bp in length
* 60921 61020: gap of 100 bp
* 61021 61885: contig of 865 bp in length
* 61886 61985: gap of 100 bp
* 61986 62792: contig of 807 bp in length
* 62793 62892: gap of 100 bp
* 62893 63744: contig of 852 bp in length
* 63745 63844: gap of 100 bp
* 63845 64693: contig of 849 bp in length
* 64694 64793: gap of 100 bp
* 64794 65645: contig of 852 bp in length
* 65646 65745: gap of 100 bp
* 65746 66624: contig of 879 bp in length
* 66625 66724: gap of 100 bp
* 66725 67550: contig of 826 bp in length
* 67551 67650: gap of 100 bp
* 67651 68502: contig of 852 bp in length
* 68503 68602: gap of 100 bp
* 68603 69435: contig of 833 bp in length

```

```

Query Match      2.6%; Score 19; DB 64; Length 151174;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 165 agaatggatattgaatt 183
|||||
Db 119794 AGAAGTGGATATTGAAGTT 119812

```

```

RESULT 15
AC026309
LOCUS AC026309 159852 bp DNA HTG 17-FEB-2001
DEFINITION Homo sapiens chromosome 3 clone RP11-256B12, WORKING DRAFT
SEQUENCE, 21 unordered pieces.
ACCESSION AC026309
VERSION AC026309.16 GI:12831263
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

```

SOURCE
ORGANISM

human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 159852)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louliseg,H., Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokoko,S.,
Ogih,M., Okwuonu,G., Oraguine,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
and Gibbs,R.

TITLE

Direct Submission

Unpublished

2 (bases 1 to 159852)

Worley,K.C.

REFERENCE
TITLE

Direct Submission

Submitted (22-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Feb 15, 2001 this sequence version replaced gi:9930647.

COMMENT

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: HAPA
Center clone name: RP11-256B12

Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy: 70% of reads
Chemistry: Dye-terminator Big Dye: 30% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 146110 bases at least Q40
Consensus quality: 151838 bases at least Q30
Consensus quality: 154679 bases at least Q20
Estimated insert size: 154344; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 3.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 21 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 22773: contig of 22773 bp in length
 22774 22873: gap of unknown length
 22874 41510: contig of 18637 bp in length
 41510 41610: gap of unknown length
 41610 54121: contig of 12511 bp in length
 54121 54221: gap of unknown length
 54221 65464: contig of 11243 bp in length
 65464 65565: gap of unknown length
 65565 74588: contig of 9023 bp in length
 74588 74688: gap of unknown length
 74688 82977: contig of 8289 bp in length
 82977 83077: gap of unknown length
 83077 92749: contig of 9673 bp in length
 92749 92850: gap of unknown length
 92850 101272: contig of 8423 bp in length
 101272 101372: gap of unknown length
 101372 109844: contig of 8472 bp in length
 109844 109944: gap of unknown length
 109944 117434: contig of 7489 bp in length
 117434 117534: gap of unknown length
 117534 121708: contig of 4174 bp in length
 121708 121808: gap of unknown length
 121808 127457: contig of 5649 bp in length
 127457 127557: gap of unknown length
 127557 132508: contig of 4951 bp in length
 132508 132608: gap of unknown length
 132608 137495: contig of 4887 bp in length
 137495 137595: gap of unknown length
 137595 141045: contig of 3451 bp in length
 141045 141145: gap of unknown length
 141145 144615: contig of 3470 bp in length
 144615 144715: gap of unknown length
 144715 147015: contig of 2300 bp in length
 147015 147115: gap of unknown length
 147115 149623: contig of 2508 bp in length
 149623 149723: gap of unknown length
 149723 152867: contig of 3144 bp in length
 152867 152967: gap of unknown length
 152967 156902: contig of 3935 bp in length
 156902 157002: gap of unknown length
 157002 157003: contig of 2850 bp in length.

FEATURES
 Source

1..159852
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /Chromosome="3"
 /clone="RP11-256B12"

BASE COUNT 45632 a 34294 c 33182 g 44726 t 2018 others
 ORIGIN

Query Match 2.6%; Score 19; DB 70; Length 159852;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 165 agaagtcgattgaagtt 183
 |||||
 Db 77555 AGAAGTCGATTTGAAGTT 77573

Search completed: August 23, 2001, 15:04:24
 Job time: 10211 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 12:17:22 ; Search time 157.23 Seconds
(without alignments)
2971.181 Million cell updates/sec

Title: US-09-654-652A-4
Perfect score: 744
Sequence: 1 atggttagcgcaaggattt.....gagatgacgaacctgctccg 744

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /cgnl_9/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /cgnl_9/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /cgnl_9/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /cgnl_9/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /cgnl_9/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /cgnl_9/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /cgnl_9/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /cgnl_9/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /cgnl_9/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /cgnl_9/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /cgnl_9/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /cgnl_9/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /cgnl_9/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /cgnl_9/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /cgnl_9/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /cgnl_9/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /cgnl_9/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /cgnl_9/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /cgnl_9/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /cgnl_9/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /cgnl_9/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /cgnl_9/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	2.6	609	AAF08911	Fusarium venenatum
2	19	2.6	1944	AAAX24386	Human CC chemokine
3	17	2.3	250	AAO76996	Human genome fragm
4	17	2.3	525	AAAX27861	Human CSR protein
5	17	2.3	824	AAZ87633	Sheep beta-lactogl
6	17	2.3	824	AAZ68328	Sheep beta-lactogl
7	17	2.3	3564	AAV69855	Human Y chromosome
8	17	2.3	3760	AAAX20279	Borrelia burgdorfe
9	17	2.3	6763	AAV79728	Beta-lactoglobulin
10	17	2.3	7494	AAV74380	Staphylococcus aur
11	17	2.3	10807	AAV03855	Sheep beta-lactogl

c 12	17	2.3	10807	18	AAV79725	Ovine beta-lactogl
c 13	17	2.3	12061	22	AAF28652	Partial DNA sequen
c 14	17	2.3	15144	18	AAV47715	Mouse bone morpho
c 15	16	2.2	33	21	AAA53077	Rat genomic DNA ve
c 16	16	2.2	147	14	AAQ38649	Encodes antifungal
c 17	16	2.2	147	15	AAQ70127	Antimicrobial Ct-A
c 18	16	2.2	303	21	AAQ09635	Human secreted pro
c 19	16	2.2	343	20	AAV87787	EST clone E113. H
c 20	16	2.2	384	21	AAQ00698	Human secreted pro
c 21	16	2.2	397	20	AAV89901	MS2 bacteriophage
c 22	16	2.2	478	21	AAQ54507	Arabidopsis thalia
c 23	16	2.2	479	21	AAQ53279	Arabidopsis thalia
c 24	16	2.2	594	19	AAV98649	DNA encoding a S.
c 25	16	2.2	634	19	AAV52656	Human native hepat
c 26	16	2.2	653	21	AAV13097	Aspergillus oryzae
c 27	16	2.2	746	21	AAV12605	Aspergillus oryzae
c 28	16	2.2	770	21	AAQ26389	Human secreted pro
c 29	16	2.2	1100	20	AAQ08328	Thermus thermophil
c 30	16	2.2	1130	21	AAQ46203	Arabidopsis thalia
c 31	16	2.2	1133	21	AAQ39316	Arabidopsis thalia
c 32	16	2.2	1200	22	AAQ54494	Botulin toxin hea
c 33	16	2.2	1365	22	AAQ88109	Human FLEXHT-40 nu
c 34	16	2.2	1386	22	AAV81363	Quorum sensing con
c 35	16	2.2	1407	19	AAV26297	Phosphate starvat
c 36	16	2.2	1431	21	AAV59945	Polynucleotide seq
c 37	16	2.2	1587	21	AAV65734	Streptococcus pneu
c 38	16	2.2	1722	18	AAV74718	Staphylococcus aur
c 39	16	2.2	1777	20	AAV84538	Human secreted pro
c 40	16	2.2	2192	19	AAV26284	S. pneumoniae deri
c 41	16	2.2	2528	21	AAV65738	Streptococcus pneu
c 42	16	2.2	2565	16	AAQ86895	A. nidulans fksA g
c 43	16	2.2	2696	20	AAV78223	Melon MEETR1 cDNA.
c 44	16	2.2	3094	16	AAQ91398	Porphyrromonas ging
c 45	16	2.2	3120	21	AAV65730	Streptococcus pneu

ALIGNMENTS

RESULT 1
AAF08911
ID AAF08911 standard; cDNA; 609 BP.
XX
AC AAF08911;
XX
DT 13-MAR-2001 (first entry)
XX
DE Fusarium venenatum EST SEQ ID NO:1434.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Fusarium venenatum.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
XX
PI (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olesen PB;
XX
XX WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells

PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 PS Claim 86; Page 924; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production of
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF1537 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 XX
 SQ Sequence 609 BP; 139 A; 197 C; 147 G; 120 T; 6 other;

Query Match 2.6%; Score 19; DB 21; Length 609;
 Best Local Similarity 100.0%; Pred. No. 4.6;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 485 tcataactgggtcaaggt 503
 |||||
 DB 235 tcataactgggtcaaggt 253

RESULT 2

AA24386
 ID AAX24386 standard; DNA; 1944 BP.

AC AAX24386;

DT 07-JUN-1999 (first entry)

XX Human CC chemokine receptor CCR8 DNA.

XX Chemokine receptor; CCR8; human; G protein coupled receptor; HIV;
 KW infection; therapy; immunomodulator; chemotaxis; apoptosis; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT CDS 258..1325
 FT /*tag= a

XX WO9906561-A2.

XX 11-FEB-1999.

XX 29-JUL-1998; 98WO-US15730.

XX 29-JUL-1997; 97US-0054094.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Alkhatib G, Bazan H, Berger EA, Bonner TI, Lautens L;
 PI Murphy PM, Tiffany HL;

XX WPI: 1899-153791/13.
 DR P-PSDB; AAW97868.
 XX
 PT New isolated chemokine receptor CCR8 - used to develop agents for
 PT modulating immune responses or agents for the prevention or
 PT treatment of HIV infection
 XX
 PS Claim 13; Fig 10A; 81pp; English.
 XX
 CC This is the nucleotide sequence of a DNA clone encoding a novel
 CC human CC chemokine receptor (see AAW97868) that has been designated
 CC CCR8. CCR8 is a G protein coupled receptor that is a necessary
 CC cofactor for infection by HIV isolates. It plays an essential role
 CC in the membrane fusion step of infection, and interacts with
 CC envelopes from all 3 classes of HIV-1. Degenerate PCR primers (see
 CC AAX24387-88) based on the predicted 3rd and 7th transmembrane domains
 CC of CXCR2 and an orphan receptor named 9-6 were used to amplify a
 CC human genomic sequence named CV6. A 1953 bp fragment of a
 CC genomic clone containing CV6 was isolated and sequenced. It
 CC included an open reading frame for a polypeptide termed CV6/TER1
 CC ORF CC chemokine receptor 6 or CCR8. The gene was mapped to
 CC human chromosome 3p22-p23 by FISH. CCR8 is expressed constitutively
 CC in monocytes and thymus, and may regulate monocyte chemotaxis and
 CC thymic cell line apoptosis. The susceptibility of target cells to
 CC HIV infection depends on cell surface expression of human CD4 and
 CC CCR8. Establishment of stable, non-human cell lines and transgenic
 CC mammals having cells that coexpress CD4 and CCR8 provides valuable
 CC tools for research on HIV infection. Antibodies which bind to
 CC CCR8, CCR8 variants, and CCR8-binding agents capable of blocking
 CC membrane fusion between HIV and target cells represent potential
 CC anti-HIV therapeutics.

XX Sequence 1944 BP; 547 A; 422 C; 410 G; 565 T; 0 other;

Query Match 2.6%; Score 19; DB 20; Length 1944;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 agaagtggatattgaagtt 183
 |||||

DB 1644 agaagtggatattgaagtt 1662

RESULT 3

AAQ76996

ID AAQ76996 standard; DNA; 250 BP.

XX AAQ76996;

XX 23-SEP-1994 (first entry)

XX Human genome fragment.

XX Brain; placenta; bone marrow; genetic analysis; gene mapping;
 KW detection; homology; human; adrenal tissue; ds.

XX Homo sapiens.

XX WO9401548-A.

XX 20-JAN-1994.

XX 13-JUL-1993; 93WO-GB01467.

XX 13-JUL-1992; 92GB-0014857.

XX (MEDIC-) MEDICAL RES COUNCIL.

XX Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;
 PI Sibson DR, Starkey M;

DR WPI; 1994-035056/04.
 XX New nucleic acid fragment encoding gene products - can be used
 PT for genetic analysis and mapping

XX Claim 1; Page 309; 616pp; English.

XX Human nucleic acid fragments, isolated from brain adrenal tissue,
 CC the placenta or bone marrow comprise any of: (A) a sequence
 CC selected from (AAQ76401-Q77613), (B) an allelic variation of a
 CC sequence as described in (A), or (C) a sequence complementary
 CC to (A) or (B).

XX Sequence 250 BP; 71 A; 63 C; 63 G; 53 T; 0 other;

Query Match 2.3%; Score 17; DB 15; Length 250;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 tccagtcacacattcatt 222
 |||||
 Db 59 tccagtcacacattcatt 75

RESULT 4

AAAX27861
 ID AAX27861 standard; DNA; 525 BP.

AC AAX27861;

XX 02-JUN-1999 (first entry)

XX Human CSR protein coding sequence fragment.

XX Cellular stress response protein; CSR1; CSR2; CSR3; human; macrophage;
 KW scavenger receptor protein; intracellular stress; arteriosclerosis;
 KW diabetic circulatory obstruction; microbial infection; ss.

XX Homo sapiens.

XX WO9909159-A1.

XX 25-FEB-1999.

XX 12-AUG-1998; 98WO-JP03602.

XX 30-JUL-1998; 98JP-0230121.

XX 13-AUG-1997; 97JP-0233396.

XX (NISR) JAPAN TOBACCO INC.

XX Nakamura Y, Tokino T;

XX WPI; 1999-181032/15.

XX Scavenger receptor proteins - for treatment and diagnosis of
 PT disorders involving cell stress

XX Example 3; Page 114-115; 175pp; Japanese.

XX This sequence represents a fragment of a DNA encoding a human cellular
 CC stress response (CSR) protein of the invention. The CSR proteins are
 CC macrophage scavenger receptor proteins. The CSR proteins can be used in
 CC the treatment, gene therapy and diagnosis of diseases in which
 CC intracellular stress is important, such as arteriosclerosis, diabetic
 CC circulatory obstruction, and microbial infection. Expression of the
 CC proteins is induced in vivo in response to intracellular stress, and
 CC inhibits cell death as a result of such stress.

XX Sequence 525 BP; 111 A; 158 C; 151 G; 104 T; 1 other;

Query Match 2.3%; Score 17; DB 20; Length 525;
 Best Local Similarity 100.0%; Pred. No. 50;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 gtcagttccatgttctt 122
 |||||
 Db 211 gtcagttccatgttctt 227

RESULT 5

AAZ87633/C

ID AAZ87633 standard; DNA; 824 BP.

XX AAZ87633;

XX 04-MAY-2000 (first entry)

XX Sheep beta-lactoglobulin (betaLG) gene fragment.

XX Transgenic bovine; transgene; milk; serum protein; industrial enzyme;
 KW infant formulation; lactoferrin; intestinal tract infection; lysozyme;
 KW iron absorption; albumin; antibacterial; iron sequestration; betaLG;
 KW beta-lactoglobulin; sheep; ds.

XX Ovis sp.

XX US6013857-A.

XX 11-JAN-2000.

XX 05-JUN-1995; 95US-0464167.

XX 16-NOV-1993; 93US-0154019.

XX 01-DEC-1989; 89US-044745.

XX 27-NOV-1990; 90US-0619131.

XX 15-JUN-1992; 92US-0898956.

XX 15-JUN-1993; 93US-0077788.

XX (PHAR-) PHARMING BV.

XX Deboer HA, Heyneker HL, Platenburg G, Krimpenfort PJA, Lee SH;

XX Pieper F, Strijker R;

XX WPI; 2000-146563/13.

XX Transgenic cattle containing transgene controlled by mammary-specific
 PT regulator, for expressing proteins in the milk, particularly human
 PT lactoferrin for infant feeding formulations

XX Example 20; Fig 24; 92pp; English.

XX The invention provides a transgenic bovine in which the somatic and germ
 CC cells contain a transgene comprising a regulatory sequence from a gene
 CC expressed in mammary glands, DNA encoding a signal sequence and DNA
 CC encoding a naturally occurring heterologous polypeptide. The transgenic
 CC bovine, or its descendants, produce milk containing the heterologous
 CC polypeptide. The transgenic bovines are used to express human milk and
 CC serum proteins or industrial enzymes, specifically for infant
 CC formulations that contain human lactoferrin for control of intestinal
 CC tract infections and to improve iron absorption, particularly when
 CC potentiated by human lysozyme. The polypeptide expressed may also be
 CC human albumin, used as a plasma extender. The polypeptide expressed in
 CC milk of the transgenic bovine requires little if any purification before
 CC human consumption and is expressed at significantly higher levels than in
 CC transgenic mice or sheep. Large polypeptides that are difficult to
 CC express in other systems can also be expressed. The present sequence
 CC represents a sheep beta-lactoglobulin gene fragment.

XX Sequence 824 BP; 138 A; 278 C; 245 G; 163 T; 0 other;

Query Match 2.3%; Score 17; DB 21; Length 824;
 Best Local Similarity 100.0%; Pred. No. 49;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 590 gctggggcaagggtgac 606
 |||||
 Db 281 GCTGGGGCAAGGTGAC 265

RESULT 6
 AAC68328/C
 ID AAC68328 standard; DNA; 824 BP.

XX AC AAC68328;

XX XX 20-FEB-2001 (first entry)

XX DE Sheep beta lactoglobulin DNA.

XX KW Lactoferrin; mammary; milk; ss..

XX OS Ovis aries.

XX XX US6140552-A.

XX PN 31-OCT-2000.

XX PD 07-JUN-1995; 95US-0476798.

XX PR 16-NOV-1993; 93US-0154019.

XX PR 01-DEC-1989; 89US-0444745.

XX PR 27-NOV-1990; 90US-0619131.

XX PR 15-JUN-1992; 92US-0898956.

XX PR 15-JUN-1993; 93US-0077788.

XX PA (PHAR-) PHARMING BV.

XX PI Strijker R, Heyneker HL, Platenburg G, Pieper F, Krimpenfort PJA;

XX PI Lee SH, Deboer HA;

XX XX WPI; 2001-040323/05.

XX DR New transgenic bovine whose mammary gland cells contain DNA encoding a

XX PT signal sequence, and a polypeptide of interest and an expression

XX PT regulatory sequence, for producing polypeptides in bovine milk -

XX XX Disclosure; Fig 24; 88pp; English.

XX CC The present invention relates to a transgenic or chimeric bovine whose

XX CC mammary gland cells contain a construct encoding a signal sequence, a

XX CC polypeptide of interest and a regulatory sequence that promotes

XX CC expression of the DNA sequence. The transgenic or chimeric bovine is

XX CC useful for producing recombinant polypeptides in milk of female

XX CC transgenic mammals. The recombinant polypeptide may be used in food

XX CC formulations, particularly in infant formula having either

XX CC nutritional or beneficial value. An infant formula containing human

XX CC lactoferrin from the transgenic bovine milk provides bacteriostatic

XX CC effect, which aids in controlling diarrhoea in newborn. Recombinant

XX CC polypeptides may also be used to supplement common diet formulations.

XX XX Sequence 824 BP; 138 A; 278 C; 245 G; 163 T; 0 other;

Query Match 2.3%; Score 17; DB 22; Length 824;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 590 gctggggcaagggtgac 606
 |||||
 Db 281 GCTGGGGCAAGGTGAC 265

RESULT 7
 AAV69855/C
 ID AAV69855 standard; DNA; 3564 BP.

XX

AC AAV69855;

DT 28-JAN-1999 (first entry)

XX DE Human Y chromosome fragment #4.

XX KW Human; Y chromosome; in situ hybridisation; probe; gender; prenatal;

XX KW forensic investigation; ss.

XX OS Homo sapiens.

XX PN US5840482-A.

XX PD 24-NOV-1998.

XX PF 10-OCT-1990; 90US-0594921.

XX PR 10-OCT-1990; 90US-0594921.

XX XX (REGC) UNIV CALIFORNIA.

XX PI Gray JW, Weier H;

XX DR WPI; 1999-034027/03.

XX PT Probes for detecting Y chromosome DNA sequence - useful for gender,

XX PT prenatal or forensic investigations

XX PS Example 1; Fig 1; 28pp; English.

XX CC The present invention describes probes for detecting a highly repeated

XX CC human Y chromosome DNA sequence. The present sequence represents a

XX CC fragment of the Y chromosome DNA to parts of which a probe could

XX CC hybridise. The probes are useful for: (1) gender determination where a

XX CC family history indicates the possibility of a sex-linked disease e.g.

XX CC haemophilia, Lesch-Nyhan syndrome, Fabry disease or Duchenne muscular

XX CC dystrophy; (2) prenatal diagnosis e.g. to enable obstetricians to

XX CC instigate treatment regimes based on gender as female foetuses have

XX CC greater survivability compared to male foetuses of the same size and

XX CC gestational age, and also to permit first trimester abortion in families

XX CC with a strong gender preference particularly in countries where

XX CC opportunities for women are very limited and where the infanticide of

XX CC female newborns is commonly practised; and (3) in forensic investigations

XX CC to detect the presence of severely decayed sperm, and to type for gender

XX CC hair and fragmentary tissue samples.

XX SQ Sequence 3564 BP; 742 A; 1081 C; 298 G; 1443 T; 0 other;

Query Match 2.3%; Score 17; DB 20; Length 3564;

Best Local Similarity 100.0%; Pred. No. 48;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 300 ctacggtctcgaatgga 316

Db 3507 CTACGGTCTCGAATGGA 3491

RESULT 8

AAX20279/C
 ID AAX20279 standard; DNA; 3760 BP.

XX AC AAX20279;

XX DT 04-MAY-1999 (first entry)

XX DE Borrelia burgdorferi polynucleotide sequence #32.

XX KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;

XX KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;

XX KW infection; diagnosis; characterisation; detection; ds.

```

OS Borrelia burgdorferi.
XX
PN WO9858943-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US12764.
XX
PP 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
XX
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
PI White OR;
XX
DR WPI; 1999-081217/07.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
XX
PS Claim 1; Page 991-993; 1128pp; English.
XX
CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
XX
SQ Sequence 3760 BP; 1477 A; 435 C; 435 G; 1412 T; 1 other;

Query Match 2.3%; Score 17; DB 20; Length 3760;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 aagggttataagtatac 515
| | | | | | | | | | | | | | | |
Db 3075 AAGCTTTATAAGTATAC 3059

RESULT 9
AAT79728/c
ID AAT79728 standard; DNA; 6763 BP.
XX
AC AAT79728;
XX
DT 11-DEC-1997 (first entry)
XX
DE Beta-lactoglobulin cloning vector pMAD6.
XX
KW Protein C; transgenic animal; sheep; rabbit; cattle; goat; milk;
KW blood clotting; anticoagulant; beta-lactoglobulin; pMAD6; vector;
KW ss.
XX
OS Chimeric Ovis aries;
OS Chimeric synthetic.
XX
PN WO9720043-A1.
XX
PD 05-JUN-1997.
XX
PP 26-NOV-1996; 96WO-US18866.
PR 13-JUN-1996; 96US-0019692.

```

```

PR 30-NOV-1995; 95US-0565074.
XX
XX (PPLT-) PPL THERAPEUTICS.
PA (ZYMO ) ZYMOGENETICS INC.
XX
XX Cottenham I, Foster DC, Garner I, Prunkard DE;
PI Sprecher CA, Temperley SM;
XX
XX WPI; 1997-310599/28.
XX
XX Production of protein C in transgenic animal - useful for high
XX quantity protein C production with therapeutic value
PT
XX
XX Example 1; Page 80-85; 99pp; English.
XX
XX This DNA sequence comprises plasmid pMAD6, a pUC18-based vector
XX containing ovine beta-lactoglobulin gene sequences (see AAT79725).
XX Nucleic acids encoding two-chain cleavage site-modified human
XX protein C (see AAT79723-24) can be cloned into the vector so that
XX the protein C gene is under the control of the beta-lactoglobulin
XX promoter. This construct can be used in claimed methods for the
XX production of human protein C in the milk of transgenic animals.
XX Modification of the protein C two-chain cleavage site improves the
XX maturation of single-chain protein C to two-chain protein C.
XX
XX Sequence 6763 BP; 1473 A; 1817 C; 1709 G; 1764 T; 0 other;
SQ

Query Match 2.3%; Score 17; DB 18; Length 6763;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 590 gctggggcaaggggtgac 606
| | | | | | | | | | | | | | | |
Db 3774 GCTGGGGCAAGGGGTGAC 3758

RESULT 10
AAV74380/c
ID AAV74380 standard; DNA; 7494 BP.
XX
AC AAV74380;
XX
DT 16-MAR-1999 (first entry)
XX
DE Staphylococcus aureus contig SEQ ID #69.
XX
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
XX
OS Staphylococcus aureus.
XX
XX
XX Key Location/Qualifiers
FT Misc_feature 901..960
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 2701..2760
FT /*tag= b
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 4501..4560
FT /*tag= c
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT

```



```

XX PF 26-NOV-1996; 96WO-US18866.
XX XX
XX PR 13-JUN-1996; 96US-0019692.
XX PR 30-NOV-1995; 95US-00565074.
XX XX
XX PA (PPLT-) PPL THERAPEUTICS.
XX PA (Zymo ) ZYMOGENETICS INC.
XX XX
XX PI Cottingham I, Foster DC, Garner I, Prunkard DE;
XX PI Sprecher CA, Temperley SM;
XX XX
XX DR WPI; 1997-310599/28.
XX PT Production of protein C in transgenic animal - useful for high
XX PT quantity protein C production with therapeutic value
XX XX
XX PS Disclosure; Page 66-74; 99pp; English.
XX XX
XX CC This DNA sequence comprises the ovine beta-lactoglobulin gene
XX CC including the promoter region. A DNA segment encompassing the
XX CC 5' flanking promoter region can be utilised in novel constructs
XX CC for the expression of human protein C in the milk of a transgenic
XX CC animal. A claimed method involves: (a) providing a DNA construct
XX CC comprising DNA encoding a secretion signal and a protein C
XX CC propeptide, operably linked to DNA encoding two-chain cleavage
XX CC site-modified protein C (see AAM25085-86), the 2 DNA sequences being
XX CC linked to elements required for protein C expression in a mammary
XX CC gland of a host female animal, such as beta-lactoglobulin gene
XX CC promoter; and (b) using the DNA construct to breed a transgenic
XX CC animal (esp. sheep, rabbit, cattle, goat) that produces protein C
XX CC in its milk, at least 90% of the protein C being in the two-chain
XX CC form.
XX XX
XX SQ Sequence 10807 BP; 2214 A; 3066 C; 2985 G; 2542 T; 0 other;

Query Match 2.3%; Score 17; DB 18; Length 10807;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 590 gctgggggcaagggtgac 606
DB 3776 GCTGGGGCAAGGGTGAC 3760
|||||
RESULT 13
AAF28652/c
ID AAF28652 standard; DNA; 12061 BP.
XX AC
XX AC AAF28652;
XX XX
XX DT 03-APR-2001 (first entry)
XX XX
XX DE Partial DNA sequence #1 of pCLYSM.
XX XX
XX KW Transgenic; milk; lysozyme; pCLYSM; human; salmon; calcitonin; ovine;
XX KW beta-lactoglobulin promoter; ss.
XX OS Unidentified.
XX XX
XX PN WO200100855-A1.
XX XX
XX PD 04-JAN-2001.
XX XX
XX PF 23-JUN-2000; 2000WO-GB02459.
XX XX
XX PR 23-JUN-1999; 99GB-0014733.
XX PR 10-AUG-1999; 99US-0147819.
XX XX
XX PA (PPLT-) PPL THERAPEUTICS SCOTLAND LTD.
XX XX
XX PI Cottingham IR, McCreath GE;

```

```

XX WPI; 2001-137958/14.
XX XX
XX PT Producing peptides for use as medicaments, nutritional additives or
XX PT research tools, comprises expressing a fusion protein, which contains
XX PT the peptide linked to a lysozyme fusion partner protein, in the milk of
XX PT a transgenic mammal -
XX XX
XX PS Example 1; Fig 2; 57pp; English.
XX XX
XX CC The present invention relates to a method for producing a peptide. The
XX CC method comprises expressing in the milk of a transgenic non-human
XX CC placental mammal a fusion protein, comprising the peptide to be expressed
XX CC linked to lysozyme (a fusion partner protein). The method is useful for
XX CC producing proteins or peptides that are useful as medicaments,
XX CC nutritional additives or research tools. The present sequence is a
XX CC partial DNA sequence of pCLYSM. pCLYSM was designed to express a human
XX CC lysozyme-salmon calcitonin fusion protein in the milk of transgenic
XX CC animals. The fusion protein allows the release of calcitonin from the end
XX CC of a linker arm fused to the lysozyme C terminal by cyanogen bromide
XX CC chemical cleavage. pCLYSM consists of the ovine beta-lactoglobulin (BLG)
XX CC promoter and 5' UTR, human lysozyme gene, peptide linker (see AAB61572),
XX CC CNBR cleavage site and salmon calcitonin coding sequence, ovine BLG 3',
XX CC UTR, polyadenylation site and flanking region, chick alpha-globin
XX CC insulator region and the pUC18 bacterial plasmid vector. The present
XX CC sequence is the DNA sequence of pCLYSM as described above, but excluding
XX CC the bacterial plasmid sequence.
XX XX
XX SQ Sequence 12061 BP; 3029 A; 2803 C; 2777 G; 3452 T; 0 other;

Query Match 2.3%; Score 17; DB 22; Length 12061;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 590 gctgggggcaagggtgac 606
DB 3792 GCTGGGGCAAGGGTGAC 3776
|||||
RESULT 14
AAT47715
ID AAT47715 standard; DNA; 15144 BP.
XX AC
XX AC AAT47715;
XX XX
XX DT 20-MAY-1997 (first entry)
XX XX
XX DE Mouse bone morphogenetic protein-2 gene and promoter region.
XX XX
XX KW Osteogenic agent; bone morphogenetic protein-2; BMP-2;
XX KW growth factor; osteoblast; promoter; osteoporosis; fracture repair;
XX KW osteoblastic metastasis; osteosclerosis; therapy; ss.
XX XX
XX OS Mus sp.
XX XX
XX PN WO9638590-A1.
XX XX
XX PD 05-DEC-1996.
XX XX
XX PF 31-MAY-1996; 96WO-US08197.
XX XX
XX PR 02-JUN-1995; 95US-0458434.
XX XX
XX PA (OSTE-) OSTEOSCREEN INC.
XX XX
XX PI Feng JQ, Ghosh-Choudhury N, Harris SE, Mundy GR;
XX XX
XX XX WPI; 1997-034396/03.
XX XX
XX PT System for identifying osteogenic agents that induce prodn. of bone
XX PT morphogenetic protein - is cell contg. reporter gene under control of
XX PT BMP gene promoter, also new promoters of BMP-2 and -4 and related

```

PT vectors and cells
 XX Example 2; Page 33-42; 76pp; English.
 XX
 CC A DNA sequence (AA747715) comprises the mouse morphogenetic protein-2
 CC (BMP-2) gene and promoter regions. It was isolated from a mouse
 CC genomic library (B6/CBA) using mouse BMP-2 cDNA as probe. Analysis
 CC of the 5' flanking sequence (see also AA747713) of the gene revealed
 CC putative response elements and transcription factor recognition
 CC sequences. The promoter region of the BMP-2 or BMP-4 (see also
 CC AA747711) gene can be linked to a reporter gene in a recombinant
 CC expression vector. Host cells transformed with the vector can be
 CC used to identify novel osteogenic agents that stimulate bone cells
 CC to produce BMP family bone growth factors. Such osteogenic agents
 CC can be used to promote bone growth and healing.
 XX
 SQ Sequence 15144 BP; 4062 A; 3380 C; 3503 G; 4199 T; 0 other;
 Query Match 2.3%; Score 17; DB 18; Length 15144;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 tggaagccctgggtag 166
 Db 8938 tggaagccctgggtag 8954
 |||||

RESULT 15
 AAA53077
 ID AAA53077 standard; DNA; 33 BP.
 XX
 AC AAA53077;
 XX
 DT 15-SEP-2000 (first entry)
 XX
 DE Rat genomic DNA vector pSP72 oligonucleotide Tag 1 SEQ ID NO:1.
 XX
 KW Genome analysis; tag; chromosomal location; integration site;
 KW insertion element; identification; polymorphism; PCR; primer; ss.
 XX
 OS Rattus sp.
 OS Synthetic.
 XX
 PN WO200024937-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 27-OCT-1999; 99WO-US25037.
 XX
 PR 28-OCT-1998; 98US-0105914.
 PR 26-OCT-1999; 99US-0427834.
 XX
 PA (STRA/) STRATHMANN M.
 XX
 PI Strathmann M;
 XX
 DR WPI; 2000-350769/30.
 XX

Search completed: August 23, 2001, 15:01:15
 Job time: 9833 sec

Query Match 2.2%; Score 16; DB 21; Length 33;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 ggaaggtgcccaggtt 381
 Db 9 ggaaggtgcccaggtt 24
 |||||

CC comparing them to a sequence database; (5) identifying a gene associated
 CC with a phenotype by assaying a distribution of polymorphisms in a
 CC population to identify a subset of the polymorphisms that associate with
 CC the phenotype and identifying the gene linked to the subset of
 CC polymorphisms or by locating the gene to within a defined genomic region
 CC and sequencing the region; (6) identifying a compound that interacts
 CC with a gene product or polypeptide or modulates expression of a gene;
 CC (7) preparing antibodies to a polypeptide; (8) preparing an array of
 CC oligonucleotides for assaying polynucleotides of known sequence; (9)
 CC producing a transgenic cell; (10) producing a cell with a mutated gene
 CC by sequencing and identifying a gene, introducing a mutation into the
 CC gene in the cell; (11) assaying gene expression; (12) identifying a
 CC bioactive polypeptide; (13) locating genome rearrangements by generating
 CC physical maps of polynucleotides and comparing them to a database of
 CC physical maps and identifying differences in landmarks to locate the
 CC genomic rearrangements; and (14) identifying a gene associated with a
 CC disease. Cells produced can be used to generate transgenic organisms
 CC which are then used to identify bioactive compounds. AAA53077 to AAA53125
 CC represent oligonucleotide sequences used in the exemplification of the
 CC present invention.

XX Sequence 33 BP; 8 A; 7 C; 13 G; 5 T; 0 other;
 SQ

PT Parallel methods of genomic analysis useful for determining
 XX polymorphisms, chromosome rearrangements and generating physical maps -
 XX Example 1; Page 91; 153pp; English.
 XX
 CC The present invention describes parallel methods of genomic analysis
 CC using polynucleotides associated with sample tags. The parallel methods
 CC are used for the following claimed procedures: (1) constructing a
 CC recombinant molecule consisting of a sequence element from a homologue
 CC of a sequenced polynucleotide joined to a vector; (2) producing a
 CC polypeptide, preferably of known sequence, from a recombinant molecule
 CC as in (1) which has been transferred to a host; (3) constructing a
 CC database of genetic information from polynucleotides sequenced using the
 CC method; (4) identifying polymorphisms by sequencing polynucleotides and

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 23, 2001, 11:45:08 ; Search time 1867.27 Seconds
(without alignments)
3766.420 Million cell updates/sec

Title: US-09-654-652A-4
Perfect score: 744
Sequence: 1 atggttagcgcaaggattt.....gagatgacgaacctgtctcg 744

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 0
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*
44: gb_est44:*
45: gb_est45:*
46: gb_est46:*
47: gb_est47:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est33:*
111: gb_est34:*
112: gb_est35:*
113: gb_est36:*
114: gb_est37:*
115: gb_est38:*
116: gb_est39:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Prd. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20	2.7	839	146	BF261375 HV_CEA000
C 2	19	2.6	517	226	AQ285087 RPC111-79
C 3	19	2.6	559	242	AZ361905 IM0106J15
C 4	19	2.6	626	258	AZ354335 Leishmani
C 5	19	2.6	636	228	AQ427471 CITBI-E1-
C 6	18	2.4	275	29	AV357168 AV357168
C 7	18	2.4	299	172	BG012371 RC2-GN028
C 8	18	2.4	332	243	AZ426939 IM0208H08
C 9	18	2.4	342	107	AU055708 AU055708
C 10	18	2.4	362	15	AI069694 TENU2738
C 11	18	2.4	371	21	AI544667 fb77d09.x
C 12	18	2.4	429	115	AW390681 CM2-ST019
C 13	18	2.4	436	235	AQ886579 HS_5544_A
C 14	18	2.4	446	137	BE006626 WHE0901_G
C 15	18	2.4	448	148	BF422269 FM1_13_D0
C 16	18	2.4	454	17	AI186540 qd35d03.x
C 17	18	2.4	461	167	BE404293 WHE1204_E
C 18	18	2.4	464	234	AQ862731 nbeb0019M
C 19	18	2.4	469	232	AQ724042 HS_5375_A
C 20	18	2.4	470	8	AA542472 fa07e08.s
C 21	18	2.4	488	5	AA308387 EST179219
C 22	18	2.4	502	167	BE422602 WHE0055_A
C 23	18	2.4	509	145	BF202269 WHE0984_F
C 24	18	2.4	511	166	BE355369 DGL_114_C
C 25	18	2.4	523	166	BE355291 DGL_114_C
C 26	18	2.4	527	146	BF259611 HVSMET001
C 27	18	2.4	551	258	TA7G12Q T. brucei
C 28	18	2.4	562	258	TA19G07P T. brucei
C 29	18	2.4	580	243	AZ426948 LM0208J08
C 30	18	2.4	586	24	A1729124 BNLGH126
C 31	18	2.4	592	122	AW961291 EST373363
C 32	18	2.4	592	234	AQ865167 nbeb0024J
C 33	18	2.4	605	105	AL036448 DKEZp564D
C 34	18	2.4	606	146	BF291791 WHE2204_A
C 35	18	2.4	607	167	BE423985 WHE0066_F
C 36	18	2.4	620	231	AQ657042 Sheared D
C 37	18	2.4	626	245	AZ519171 RPC1-11-7
C 38	18	2.4	638	233	AQ754786 HS_3166_B
C 39	18	2.4	644	166	BE356747 DGL_12_B0
C 40	18	2.4	696	16	AI110205 GH09761.5
C 41	18	2.4	702	223	AQ013977 RPC111-25
C 42	18	2.4	753	141	BE885533 601508727
C 43	18	2.4	822	153	BG418473 HVSMET002
C 44	18	2.4	825	233	AQ747536 HS_5536_A
C 45	18	2.4	867	146	BF267575 HV_CEA001

ALIGNMENTS

RESULT 1	BF261375/c	839 bp	EST	09-MAR-2001
LOCUS	HV_Cea0001B22f	Hordeum vulgare seedling green leaf EST library		
DEFINITION	HVCNDA0004 (Erysiphe infected & control) Hordeum vulgare cDNA clone			
ACCESSION	BF261375			
VERSION	BF261375.2	GI:13259773		
KEYWORDS	barley.			
SOURCE	Hordeum vulgare			
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae			
	; Triticeae; Hordeum.			
REFERENCE	1 (bases 1 to 839)			
AUTHORS	Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu			
	Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo			

TITLE	Development of a genetically and physically anchored EST resource for barley genomics			
JOURNAL	Unpublished (2000)			
COMMENT	On Nov 17, 2000 this sequence version replaced gi:11192362.			
	Contact: Wing RA			
	Clemson University Genomics Institute			
	Clemson University			
	100 Jordan Hall, Clemson, SC 29634, USA			
	Tel: 864 656 7288			
	Fax: 864 656 4293			
	Email: rwing@clemson.edu			
	Seq primer: AATTAACCTCTACTAAAGGG			
	High quality sequence stop: 719.			
FEATURES	Location/Qualifiers			
	1..839			
	/organism="Hordeum vulgare"			
	/cultivar="C116155 (Mial3)"			
	/db_xref="taxon:4513"			
	/clone_lib="Hordeum vulgare seedling green leaf EST			
	library HVCNDA0004 (Erysiphe infected & control)"			
	/tissue_type="seedling green leaf"			
	/lab_host="TJCI21"			
	/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"			
	BASE COUNT 159 a 247 c 225 g 208 t			
ORIGIN				
	Query Match 2.7%; Score 20; DB 146; Length 839;			
	Best Local Similarity 100.0%; Pred. No. 5.3;			
	Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	452	agtcgatgaatcaagcctt 471		
Db	739	AGTTCGATGAATCAAGCCTT 720		
RESULT 2				
LOCUS	AQ285087	517 bp	DNA	GSS 04-MAY-1999
DEFINITION	RPC111-79N17.TJ RPC1-11 Homo sapiens genomic clone RPC1-11-79N17,			
	DNA sequence.			
ACCESSION	AQ285087			
VERSION	AQ285087.1	GI:3911327		
KEYWORDS	GSS.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 517)			
AUTHORS	Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,			
	Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.			
TITLE	Use of human BAC End Sequences for Sequence-Ready Map Building			
JOURNAL	Unpublished (1998)			
COMMENT	Other GSSs: RPC111-79N17.TV			
	Contact: Mark Adams			
	Department of Eukaryotic Genomics			
	The Institute for Genomic Research			
	9712 Medical Center Dr., Rockville, MD 20850, USA			
	Tel: 301 838 0200			
	Fax: 301 838 0208			
	Email: mdadams@tigr.org			
	Clones are derived from the human BAC library RPC1-11. For BAC			
	library availability, please contact pieter de Jong			
	(pieter@dejong.med.buffalo.edu). Clones may be purchased from			
	BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from			
	Research Genetics (info@resgen.com). BAC end search page:			
	http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html			
	Seq primer: SP6			
	Class: BAC ends.			
FEATURES	Location/Qualifiers			

source

1. .517
 /organism="Homo sapiens"
 /db_xref="GDB:7530280"
 /db_xref="taxon:9606"
 /clone="RPCI-11-79N17"
 /clone_lib="RPCI-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
 RPC111 Human Male BAC Library" 1 others
 BASE COUNT 168 a 97 c 54 g 197 t
 ORIGIN

Query Match 2.6%; Score 19; DB 226; Length 517;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 tcagttccatgttcctcta 125
 |||||
 Db 470 TCAGTTCATGTTCCCTCA 488

RESULT 3

AZ361905/c
 LOCUS 559 bp DNA 02-OCT-2000
 DEFINITION 1M0106J15R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0106J15 R, DNA sequence.

ACCESSION AZ361905
 VERSION
 KEYWORDS
 SOURCE

ORGANISM

Mus musculus
 house mouse.
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss R.

TITLE
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts

JOURNAL

COMMENT
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0106 row: J column: 15

Seq primer: CACACGAGGAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 559.

FEATURES

source

1. 559
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0106J15"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PWD42 (gi|4732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 177 a 103 c 114 g 165 t
 ORIGIN

Query Match 2.6%; Score 19; DB 242; Length 559;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 385 aacttgacaggtacacagg 403
 |||||
 Db 423 AACTTGACAGGTACACAGG 405

RESULT 4

L6474AX 626 bp DNA GSS 29-APR-2000
 LOCUS Leishmania major Friedlin cosmid L6474.1 t3Hyg end-sequence,
 genomic survey sequence.

ACCESSION AL354335

VERSION AL354335.1 GI:7671978

KEYWORDS

SOURCE

ORGANISM

Leishmania major.
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.

REFERENCE

AUTHORS
 Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and
 Smith,D.F.

TITLE
 A physical map of the Leishmania major Friedlin genome

JOURNAL

MEDLINE

98146435

REFERENCE

AUTHORS
 2 (bases 1 to 626)
 Taylor,R.G., Huckle,E.E.J., Ivens,A.C., Rajandream,M.A. and
 Barrell,B.G.

TITLE
 Direct Submission

JOURNAL
 Submitted (28-APR-2000) Leishmania major Friedlin genome sequencing
 project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
 Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
 alicat@sanger.ac.uk

COMMENT
 see http://www.ebi.ac.uk/parasites/leish.html

Details of Leishmania sequencing at the Sanger Centre are available
 at http://www.sanger.ac.uk/Projects/L_major/
 The cLHYg t3Hyg primer sequence can be obtained from acc. no.
 U59231.

FEATURES

source

1. 626
 Location/Qualifiers
 /organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /clone="cosmid L6474.1"
 BASE COUNT 96 a 213 c 186 g 129 t 2 others
 ORIGIN

Query Match 2.6%; Score 19; DB 258; Length 626;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 523 cagggcgaagcgccagcg 541
 |||||
 Db 79 CAGGGCGAAGCGGCAGCG 97

RESULT	5	
AQ427471/c		
LOCUS	636 bp	DNA
DEFINITION	GSS 24-MAR-1999 CITBI-El-2574N14.TF CITBI-El Homo sapiens genomic clone 2574N14, DNA sequence.	
ACCESSION	AQ427471	
VERSION	AQ427471.1	GI:4500376
KEYWORDS	GSS.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 635) Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and Venter,J.C.	
TITLE	Use of BAC End Sequences from CalTech Libraries for Sequence-Ready Map Building	
JOURNAL	Unpublished (1997)	
COMMENT	Other GSSs: CITBI-El-2574N14.TR Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208 Email: hbs@tigr.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html . Seq primer: M13-21 Class: BAC ends.	

```

FEATURES
  source
    Location/Qualifiers
      1..636
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_="2574N14"
        /clone_lib="CITBI-E1"
        /sex="male"
        /cell_type="sperm"
        /note="Vector: pBelOAc11; Site_1: EcoRI; Site_2: EcoRI;
          Caltech Human BAC Library D"
      278 a 126 c 117 q 115 t
BASE COUNT

```

Query Match	2.8%;	Score 19;	DB 228;	Length 636;
Best Local Similarity	100.0%;	Pred. No. 19;		
Matches 19;	Conservative	0;	Mismatches 0;	Indels 0;
QY	111	ttccatgttctctaccag	129	
Db	494	ttccatgttctctaccag	476	

RESULT	6			
AV357168/c				
LOCUS		275 bp	EST	13-NOV-1999
DEFINITION	AV357168	RIKEN full-length enriched, in vitro fertilized eggs	Mus	
	musculus	cDNA clone 742042F18	3', mRNA	sequence..
ACCESSION	AV357168			
VERSION	AV357168.1	GI:6404170		
KEYWORDS	EST.			

SOURCE.
ORGANISM
mouse musculus
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 275)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,

Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Suganara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateo, M., Tomimaga, N., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshikawa, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. Riken Mouse ESTs (Konno, H. et al., 1999)

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1980, Vol. 83, No. 1, pp. 1-10
2. The Impact of Technology on Education	Journal of Educational Research	1980, Vol. 83, No. 2, pp. 11-20
3. The Importance of Parental Involvement	Journal of Educational Research	1980, Vol. 83, No. 3, pp. 21-30
4. The Effect of Teacher Expectations on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 4, pp. 31-40
5. The Role of the School in the Community	Journal of Educational Research	1980, Vol. 83, No. 5, pp. 41-50
6. The Impact of Socioeconomic Status on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 6, pp. 51-60
7. The Importance of Teacher Education	Journal of Educational Research	1980, Vol. 83, No. 7, pp. 61-70
8. The Effect of Teacher Attitudes on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 8, pp. 71-80
9. The Role of the School in the Community	Journal of Educational Research	1980, Vol. 83, No. 9, pp. 81-90
10. The Impact of Socioeconomic Status on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 10, pp. 91-100

unpublished (1999)
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098

Fax: +81 430 50 5020
 Email: genome-resortc.riken.go.jp,
 URL: <http://genome.rtc.riken.go.jp/>
 Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,
 Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki
 Y.

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 95 (7), 3455-3460 (1998)

Itô, M., Kutsuna, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y. Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES	SOURCE
----------	--------

```

FEATURES
source
.
```

prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTTTN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATCTCGAGTTATTAATTAATCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLIC I. Cloning sites, 5' end: SalI; 3' end: BamHI*
79 a 55 c 39 a 102 t

BASE COUNT
ORIGIN

Query Match	2.4%	Score 18;	DB 29;	Length 275;
Best Local Similarity	100.0%;	Pred. No. 67;		
Matches 18:	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

Qy	151	ggaagccctgggtaga	168
Db	20	GGAAGCCCTGGGTAGA	3

Db 20 GGAAGGCCCTGGGTAGAA 3

```

RESULT 7
BG012371      299 bp      mRNA      EST      24-JAN-2001
LOCUS      RC2-GN0280-131200-012-e02 GN0280 Homo sapiens cDNA, mRNA sequence.
DEFINITION
ACCESSION      BG012371
VERSION      BG012371.1 GI:12461495
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 299)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC2&t2=RC2-GN0280-
131200-012-e02&t3=2000-12-13&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 38
High quality sequence stop: 299.
FEATURES
source
1..299
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0280"
/dev_stage="Adult"
/note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the puc 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT      59 a      98 c      62 g      80 t
ORIGIN
1 Query Match      2.4%; Score 18; DB 172; Length 299;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 367 gaagtgccagggtttcc 384
|||||
DB 54 GAAGTGGCCAGGTTCC 71

RESULT 8
A2426939/c
LOCUS      A2426939      332 bp      DNA      GSS      03-OCT-2000
DEFINITION      1M0208H08R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0208H08 R, DNA sequence.
ACCESSION      A2426939
VERSION      A2426939
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus

```

```

REFERENCE
AUTHORS
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0208 row: H column: 08
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 332.
Location/Qualifiers
1..332
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0208H08"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD22nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD22 (g14732114[gblAF129072.1]), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      82 a      88 c      71 g      91 t
ORIGIN
1 Query Match      2.4%; Score 18; DB 243; Length 332;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 aggcctggtagaagt 171
|||||
DB 49 AGCCCTGGTAGAGTG 32

RESULT 9
A0055708
LOCUS      A0055708      342 bp      mRNA      EST      29-APR-1999
DEFINITION      Oryza sativa mature leaf Nipponbare Oryza sativa CDNA
clone S20009_1A, mRNA sequence.
ACCESSION      A0055708
VERSION      A0055708.1 GI:4714591
KEYWORDS      EST.
SOURCE      Oryza sativa.

```

```

ORGANISM
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoidae; Oryzeae; Oryza.
REFERENCE
AUTHORS
Yamamoto,K. and Sasaki,T.
TITLE
Rice cDNA from mature leaf
JOURNAL
Unpublished (1999)
COMMENT
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'

FEATURES
source
1. .342
Location/Qualifiers
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="520009.1A"
/clone_lib="Oryza sativa mature leaf Nipponbare"
/tissue_type="mature leaf"
BASE COUNT
111 a 84 c 66 g 81 t
ORIGIN
|||||
Query Match 2.4%; Score 18; DB 107; Length 342;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 agtccaacattaccg 226
|||||
Db 146 AGTCCACATCATTCACG 163

RESULT 10
AI069694/c 362 bp mRNA EST 21-SEP-2000
LOCUS
TENU2738 T. cruzi epimastigote normalized cDNA library Trypanosoma
DEFINITION
cruzi cDNA clone 26p14 5', mRNA sequence.
ACCESSION
AI069694
VERSION
AI069694.1 GI:3392669
KEYWORDS
EST.
SOURCE
Trypanosoma cruzi.
ORGANISM
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE
AUTHORS
Porcel,B.M., Tran,A.-N., Tammi,M., Nyarady,Z., Rydaker,M., Urmenyi
T.P., Rondinelli,E., Pettersson,U., Andersson,B. and Aslund,L.
TITLE
Gene survey of the pathogenic protozoan Trypanosoma cruzi
JOURNAL
Genome Res. 10 (8), 1103-1107 (2000)
MEDLINE
20414748
COMMENT
Contact: Aslund L
Department of Medical Genetics
Uppsala University
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden
Tel: 46 18 471 45 85
Fax: 46 18 52 68 49
Email: lena.aslund@medgen.uu.se
Seq primer: T7
High quality sequence stop: 362.
Location/Qualifiers
1. .362
/organism="Trypanosoma cruzi"
/strain="Cl-Brenner"
/db_xref="taxon:5693"
/clone="26p14"
/clone_lib="T. cruzi epimastigote normalized cDNA Library"

FEATURES
source
1. .362
Location/Qualifiers
/organism="Trypanosoma cruzi"
/strain="Cl-Brenner"
/db_xref="taxon:5693"
/clone="26p14"
/clone_lib="T. cruzi epimastigote normalized cDNA Library"

```

```

/cell_type="epimastigote"
/notes="cDNA library constructed with oligo dT primed
epimastigote mRNA and cloned in pT7318D phagemid with
modified polylinker (PHARMACIA)"
BASE COUNT
115 a 47 c 115 g 85 t
ORIGIN
|||||
Query Match 2.4%; Score 18; DB 15; Length 362;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 cgttagaagaagtccagt 58
|||||
Db 195 CGTTAGAAGATTCAGT 178

RESULT 11
AI544667 371 bp mRNA EST 22-MAR-1999
LOCUS
fb77d09.x1 zebrafish WashU MPIMG EST.Danio rerio cDNA 3', mRNA
DEFINITION
sequence.
ACCESSION
AI544667
VERSION
AI544667.1 GI:4462040
KEYWORDS
EST.
SOURCE
zebrafish.
ORGANISM
Danio rerio
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE
AUTHORS
Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
TITLE
WashU Zebrafish EST Project 1998
JOURNAL
Unpublished (1998)
COMMENT
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrfish@watson.wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
www.rzpd.de)
zebrafish identity (p-value greater than 1e-99) found to:
gi1289407|gb|AA542472|AA542472 fa07e08.sl zebrafish ICRFzfls Danio
rerio cDNA
Seq primer: T7 ET from Amersham
High quality sequence stop: 347.
Location/Qualifiers
1. .371
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="zebrafish WashU MPIMG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XLI-blue MRF"
/notes="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'pgactagttctagatcgagcgccgctttttttttttt3'];
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed

```

by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

BASE COUNT 90 a 94 c 65 g 122 t
ORIGIN

Query Match 2.4%; Score 18; DB 21; Length 371;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 gtccagttccattcttc 123
Db 196 gtccagttccattcttc 213

RESULT 12
LOCUS AW390681/c
DEFINITION CM2-ST0194-181099-021-g07 ST0194 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW390681
VERSION AW390681.1 GI:6895340
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 429)
AUTHORS HCGP <http://www.ludwig.org.br/ORESTES>,
TITLE The FAPESP/LICR Human Cancer Genome Project.
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&t2=CM2-ST0194-181099-021-g07&t3=1999-10-18&t4=1>)
Seq primer: puc 18 forward
High quality sequence stop: 429.

FEATURES
source
1. 429
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ST0194"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 96 a 117 c 107 g 109 t
ORIGIN

Query Match 2.4%; Score 18; DB 115; Length 429;
Best Local Similarity 100.0%; Pred. No. 69;

Matches
QY 151 ggaagccctggtagaa 168
Db 232 GGAAGGCCCTGGGTAGAA 215

RESULT 13
LOCUS AQ886579/c
DEFINITION HS_5544_A1_B05_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=9312 Col=9 Row=C, DNA sequence.
ACCESSION AQ886579
VERSION AQ886579.1 GI:6318046
KEYWORDS GSS.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 436)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering.bac.htm>) or from Resear h Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 9312 row: C column: 9
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 436.

FEATURES
source
1. 436
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=9312 Col=9 Row=C"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

BASE COUNT 97 a 91 c 86 g 160 t
ORIGIN

Query Match 2.4%; Score 18; DB 235; Length 436;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 201 cagttccagtcacacat 218
Db 343 CAGTTTCCAGTCCAACAT 326

RESULT 14
LOCUS BE606626/c
DEFINITION WHE0901_G12_M232S Wheat 5-15 DAP spike cDNA library Triticum

Query Match 2.4%; Score 18; DB 235; Length 436;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 201 cagttccagtcacacat 218
Db 343 CAGTTTCCAGTCCAACAT 326

RESULT 14
LOCUS BE606626/c
DEFINITION WHE0901_G12_M232S Wheat 5-15 DAP spike cDNA library Triticum

accession B506626
 VERSION B506626.1 GI:9883790
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum

REFERENCE 1 (bases 1 to 446)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.

AUTHORS Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han, P.S., Haia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L. and Tong, J.C.
 TITLE The structure and function of the expressed portion of the wheat genomes - 5-15 DAP spike cDNA library
 JOURNAL Unpublished (2000)
 COMMENT Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20.
 Seq primer: Stratagene SK primer.
 FEATURES
 source
 1..446
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone_lib="WHE0901_G12_M23"
 /clone_lib="Wheat 5-15 DAP spike cDNA library"
 /tissue_type="Spike"
 /dev_stage="Adult plant"
 /lab_host="E. coli SOLR"
 /notes="vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse. Spikes at 5, 10 and 15 DAP were harvested, total RNA and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the T3 Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT 143 a 96 c 111 g 96 t
 ORIGIN

Query Match 2.4%; Score 18; DB 137; Length 446;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 476 ttttcaggttcataact 493
 |||||
 Db 357 TTTTCCAGTTCATCAACT 340

RESULT 15
 BF422269/c
 LOCUS
 DEFINITION FMI_13_D07_b1_A003 Floral-Induced Meristem 1 (FMI) Sorghum propinquum cDNA, mRNA sequence.
 ACCESSION BF422269
 VERSION BF422269.1 GI:11410258
 KEYWORDS EST.
 SOURCE Sorghum propinquum.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
 REFERENCE 1 (bases 1 to 448)

AUTHORS
 TITLE
 JOURNAL
 COMMENT

Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M., Marsala, C. and Pratt, L.H.
 An EST database from Sorghum: floral-induced meristems
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: JEN REV
 High quality sequence stop: 379
 POLYA-No.

FEATURES
 source

Location/Qualifiers
 1..448
 /organism="Sorghum propinquum"
 /db_xref="taxon:132711"
 /clone_lib="Floral-Induced Meristem 1 (FMI)"
 /note="Organ: Floral-induced meristems; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested. The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 142 a 100 c 114 g 92 t
 ORIGIN

Query Match 2.4%; Score 18; DB 148; Length 448;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 476 ttttcaggttcataact 493
 |||||
 Db 367 TTTTCCAGTTCATCAACT 350

Search completed: August 23, 2001, 14:09:06
 Job time: 8638 sec

